Speech across Species

On the mechanistic fundamentals of vocal production and perception

Verena R. Ohms

Ohms, Verena Regina

Speech across Species
On the mechanistic fundamentals of vocal production and perception

Dissertation Leiden University

ISBN/EAN: 978-94-90858-06-3

An electronic version of this thesis in Adobe PDF-format is available at:
https://openaccess.leidenuniv.nl/dspace/

Printed by Mostert & Van Onderen!, Leiden

Cover design and photographs by Verena Ohms and Ion Chih

Speech across Species

On the mechanistic fundamentals of vocal production and perception

PROEFSCHRIFT

ter verkrijging van

de graad van Doctor aan de Universiteit Leiden,

op gezag van Rector Magnificus Prof. Mr. P. F. van der Heijden,

volgens besluit van het College voor Promoties

te verdedigen op dinsdag 3 mei 2011

klokke 16:15 uur

door

Verena Regina Ohms

geboren te Viersen, Duitsland
in 1983

Promotiecommissie

Promotor:	Prof. Dr. Carel J. ten Cate
Copromotor:	Dr. Gabriël J. L. Beckers (Max-Planck-Institute for Ornithology, Seewiesen, Germany)
Overige leden:	Dr. Clara C. Levelt
	Prof. Dr. Annemie van der Linden (University of Antwerp, Belgium)
	Prof. Dr. Michael K. Richardson
	Prof. Dr. Constance Scharff (Free University of Berlin, Germany)
	Prof. Dr. Herman P. Spaink

This work was supported by the Research Council for Earth and Life Sciences (ALW, grant number 815.02.011) with financial aid from the Netherlands Organization for Scientific Research (NWO).

Contents

1

General introduction, thesis overview and discussion

Studying the evolution of speech

Human language constitutes one of the most complex behaviours known to date. It allows us to build and communicate an infinite number of conceptual structures independent of modality (Fitch 2000; Brenowitz *et al.* 2010). The origin of language is unclear and there is much debate about its evolution and the particular properties that make human language unique (e.g. Hauser *et al.* 2002; Dunbar 2003; Castro *et al.* 2004; Pinker & Jackendoff 2005; Anderson 2008; Pinker 2010; Fitch 2010).

Speech on the other hand describes the actual physical phenomenon which is used to convey language. It consists of a limited number of meaningless sounds which can be combined to form a potentially infinite set of meaningful larger units (Liberman & Whalen 2000). As such speech and the mechanisms underlying its production and perception can be subjected to acoustic, physiological, anatomical and neurobiological studies. Unfortunately, such studies reveal little about the evolution of speech. Also, the fossil record of structures involved in speech production is, if at all existent, inconclusive (Fitch 2000; Ghazanfar & Rendall 2008; Fitch 2010) and therefore insufficient to reliably trace speech evolution. However, studying vocal communication in other species enables us to detect mechanisms which have evolved convergently and thus can help identify selection pressures or intermediate steps that might have caused the emergence of these mechanisms (Hauser & Fitch 2003; Jarvis 2004). This comparative approach is one of the prime methods of a young research area referred to as biolinguistics (Fitch 2010).

Insights from the comparative approach

In recent years an increasing number of studies have addressed possible similarities between human and animal vocal communication. Being a learned behaviour is one of the core properties of human speech, but vocal learning is rare in the animal kingdom (Janik & Slater 1997). Among mammals it has been found, besides in humans, only in a few distantly related groups including seals, cetaceans, bats and elephants (Janik & Slater 1997; Poole *et al.* 2005) whereas in our closest relatives, the great apes, or other primates for that matter, vocal learning seems to be largely absent (Fitch 2000).

However, vocal learning has also been demonstrated in three orders of birds, namely songbirds (Marler 1976), parrots (reviewed by Pepperberg 2010) and hummingbirds (Baptista & Schuchmann 1990). Additionally, neuroscientific studies suggest that special brain pathways for vocal learning, one posterior and one anterior, are present in all three groups of vocal learning birds and humans, but not in vocal non-learning birds or mammals (Jarvis 2004).

Moreover, several other parallels between human speech and birdsong have established birdsong as the closest animal analogue to human speech that exists and therefore as an excellent model system to study the underlying mechanisms of speech production and perception (Bolhuis *et al.* 2010).

Both humans and songbirds exhibit a sensitive period early in life during which vocal learning is facilitated (White 2001). Auditory feedback by hearing others and themselves is crucial in order to develop normal vocalizations, initially during a sensory learning phase and later during sensorimotor learning (Doupe & Kuhl 1999). It has also been hypothesized that in humans as well as birds innate predispositions and biases guide which sounds will be learned (Whaling *et al.* 1997; Braaten & Reynolds 1999). However, evidence for this is more readily available in songbirds because they can be subjected to cross-fostering (Clayton 1989) and isolation studies (Braaten & Reynolds 1999) which cannot be conducted in humans for apparent ethical reasons.

Recently, thanks to the development of new molecular techniques, another component came into focus: the genetic basis for vocal communication (Bolhuis *et al.* 2010). Mutations in the gene coding for the transcription factor FOXP2, for instance, are associated with a speech disorder in humans, manifested by impaired motor control of orofacial movements and deficits in several aspects of language processing (Lai *et al.* 2001). Interestingly, the same gene when down-regulated in Area X of juvenile zebra finch brains causes these birds to inaccurately copy tutor songs by omitting some song elements and showing more variability between song repetitions (Haesler *et al.* 2007).

The parallels described so far are mainly concentrating on vocal development and learning. However, a growing body of evidence indicates that there are also similarities between human speech and birdsong with respect to vocal production and perception. Both human speech and birdsong are produced by a primary sound source and subsequently filtered in the vocal tract (Fant 1960; Nowicki 1987), but currently vocal tract filtering is less well understood in (song)birds than it is in humans. Regarding vocal perception, which has been extensively studied in humans, it is still unclear whether specialized perceptual abilities are necessary for speech perception and whether these abilities occur in songbirds too. This thesis aims to address both vocal production and perception mechanisms in (song)birds compared to humans in order to shed more light on the similarities and differences of these mechanisms.

Formants and their relevance in vocal communication

Human speech is characterized by a broad frequency spectrum. Voiced speech sounds are produced by the vibrations of the vocal folds in the larynx, generating a fundamental frequency with harmonic overtones. This sound is filtered while traveling through the vocal tract, consisting of the pharyngeal, oral and nasal cavities. Depending on the dimensions of these cavities some frequencies within the broadband spectrum are amplified whereas others are attenuated (Titze 2000). Amplified frequencies appear as dark bands on spectrograms and are referred to as 'vocal tract resonances' or 'formants' (Fig. 1.1). It is important to notice that formants are independent of the sound source and that they are rapidly modulated during speech by moving the articulators such as tongue, lips and soft palate (Fant 1960). Formants are especially relevant in the production and perception of vowels (Ladefoged 2006). The difference between the words 'beat' and 'bit' is based on different formant values, primarily regarding the two lowest formants F1 and F2 which represent resonances of the pharynx and mouth cavity respectively.

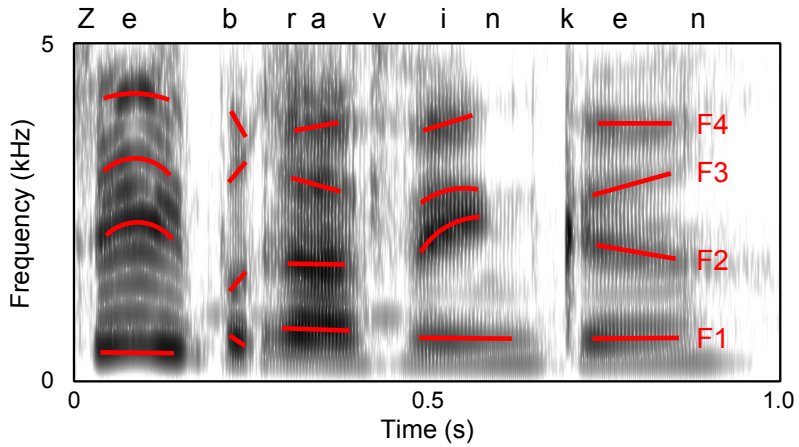

Figure 1.1 Spectrogram of human speech

This figure shows a spectrogram of a female voice saying *'zebravinken'* (zebra finches). Formant frequencies are highlighted in red. kHz, kilohertz; s, seconds; F1, first formant; F2, second formant; F3, third formant; F4, fourth formant.

During the ontogeny of modern humans the larynx descends, enabling the tongue to move both vertically and horizontally thereby allowing the production of a wide variety of formant patterns by shaping the vocal tract in numerous different ways (Lieberman *et al.* 1969). It is commonly believed that this anatomical configuration of the human vocal tract, together with the loss of laryngeal air sacs, was a necessary prerequisite for the evolution of speech (Fitch 2000). However, the idea that the evolution of speech was a driving force in the evolution of this configuration has been questioned by the discovery that several other species exhibit a descended larynx too.

It has been shown, for instance, that males of both red and fallow deer lower their larynges while roaring. This presumably serves to elongate the vocal tract causing it to resonate at lower frequencies which could make these animals sound bigger (Fitch & Reby 2001; Reby *et al.* 2005). This 'size exaggeration hypothesis' gains further support from observations of trumpet birds which exhibit elongated tracheas assumed to lower the pitch of their vocalizations (Clench 1978; Fitch 1999). At the same time this means that the descent of the larynx has been shaped by sexual selection and provided a pre-adaptation for speech evolution (Fitch 2000).

Recent evidence from comparative MRI studies also suggests that the descent of the larynx evolved before the human and chimpanzee lineages separated (Nishimura *et al.* 2006). The authors speculate that facial flattening, instead of lowering the larynx, enabled the typical configuration of the vocal tract and hence played a more important role in the evolution of speech.

However, less is known so far about the production of vocal tract resonances in songbirds. The vocal organ of songbirds, the syrinx, is much more complex than the human larynx. In Oscine songbirds two sets of vibrating labia, located at the cranial end of each bronchus (Goller & Larsen 1997), are involved in vocal production. Each of these sets can be controlled independently (Suthers 1990) and enables the birds to sing with two voices simultaneously or switch between both sets of labia while singing (Suthers 1990; Suthers *et al.* 1994; Suthers *et al.* 2004; Zollinger & Suthers 2004). This complexity has initially led to the hypothesis that acoustic variation in birdsong arises at the sound source and that, contrary to human speech production, vocal tract filtering only plays a minor role (Greenewalt 1968). Newer studies, however, suggest that there might be more parallels in human and avian sound production than originally assumed. Therefore one of the main objectives of this thesis is to identify potential articulators involved in vocal tract filtering and to evaluate their effects on sound production.

The second main objective concerns vocal perception mechanisms. Humans are highly sensitive to speech sound variation and formant patterns, but the question whether this sensitivity coevolved with speech production and is therefore a uniquely human trait or whether it is based on general auditory processing mechanisms remains highly controversial (e.g. Lieberman 1975; Kuhl & Miller 1975, 1978; Pinker & Jackendoff 2005). Some of the perception mechanisms which scientists initially claimed to occur only in humans, like the categorical perception of speech sounds, are shared with a number of different species including both birds (Kluender *et al.* 1987; Dooling & Brown 1990) and mammals (Kuhl & Miller 1975, 1978; Kuhl & Padden 1982). Another significant aspect of speech perception regards our ability to recognize words regardless of speaker identity. We can distinguish words that closely resemble each other by extracting relevant acoustic information while at the same time ignoring speaker-dependent variation. It still has to be tested if this property is uniquely human implying that it has evolved specifically for human speech or if it is, like categorical perception, based on general processing mechanisms of the auditory system.

Thesis overview

This thesis documents four studies of which two are dealing with the production and the other two with the perception mechanisms of vocal tract resonances by birds. I primarily used zebra finches in my experiments because they are the most widely used model species for studies on vocal learning, development and perception while relatively little is known about vocal production in this species. I also chose monk parakeets as another species to study vocal tract filtering since there are indications for significant differences regarding vocal production between parrots and songbirds (e.g. Beckers *et al.* 2004).

In **chapter 2** I describe an experiment which was designed to identify potential vocal tract articulators in zebra finches and to evaluate their significance on vocal tract filtering in this species. First we obtained cineradiographic movies of singing zebra finches, for which the analyses revealed beak gape and the expansion of the oropharyngeal-esophageal cavity (OEC) to be the main articulators involved in vocal production. These results are in line with earlier studies that found positive correlations between beak gape and frequency patterns in several songbird species including zebra finches (Westneat *et al.* 1993; Podos *et al.* 2004; Williams 2001). More interesting is the observation that zebra finches expand their OEC substantially while singing. This has first been demonstrated in northern cardinals (Riede *et al.* 2006) and subsequently in white-throated sparrows (Riede & Suthers 2009). Both of these species, however,

produce rather simple, pure-tone songs with little energy in higher harmonics and it has been hypothesized that OEC expansion tracks the fundamental frequency. Zebra finches on the other hand produce songs consisting of many different element types. Most of these elements are broad-band and exhibit a rich frequency spectrum including harmonic stacks with varying amplitude patterns. Due to the complexity of zebra finch song it is difficult to establish clear relationships between articulator configurations and sound patterns. Nevertheless, when experimentally manipulating beak gape and OEC expansion in the second part of the study, we found a downwards shift in peak frequency with increasing OEC expansion as well as an amplitude increase especially around 1.5 and 4.5 kHz. Beak gape on the other hand seems to emphasize frequencies around 5 kHz and above. These results demonstrate that the upper vocal tract, especially beak gape and OEC expansion, plays a major role in resonance filtering of zebra finch song resulting in elaborate note types exhibiting formant like patterns.

Parrots are another group of birds that produce complex broad-band sounds and they are very well known for their sophisticated ability to imitate human speech. In contrast to songbirds parrots have a prominent tongue with many intrinsic muscles and a fleshy, flexible surface which resembles the human tongue (Homberger 1986). This observation has motivated the hypothesis that tongue movements play a more important role in vocal production in this group of birds compared to songbirds (Patterson & Pepperberg 1994) and observations of a speech-imitating parrot (Warren *et al.* 1996) as well as experimental manipulations of tongue position (Beckers *et al.* 2004) support this claim. However, to date no direct observations of tongue movements in naturally vocalizing parrots exist, nor is it not known what other articulators are involved in vocal production in parrots.

In **chapter 3** of this thesis I therefore address this question by employing cinematographic imaging of naturally vocalizing monk parakeets. On the videos we could identify three main articulatory movements: beak opening, tongue height changes and tracheal shortening. Although earlier studies already indicated the significance of tongue movements, they found main effects in the front/back dimension of tongue position while the parakeets in our study primarily manipulated tongue height. From the nine different vocalization types produced by adult monk parakeets (Martella & Bucher 1990) the birds in our study uttered only three. This leaves the possibility that tongue movements in the front/back dimension are of significant importance in some of the other vocalizations that we could not record. Yet, in greeting calls which exhibit formant changes and which are included in our analysis, manipulations of tongue position in the vertical dimension

seem more prominent than changes in the horizontal plane. Interestingly, we also found evidence for tracheal shortening whereas an earlier study on zebra finches concluded that tracheal length changes are too small to affect vocal production in that species (Daley & Goller 2004). Furthermore we found significant positive correlations between sound amplitude and magnitude of articulator movements in greeting calls and chatter sounds for beak opening, tongue height and tracheal shortening for some of the birds. Since modulations of the fundamental frequency (F0) are very fast in monk parakeet contact calls while articulator movements are comparatively slow it is likely that changes in F0 are generated at the sound source. Formant patterns as occurring in greeting calls, however, are probably the result of the vocal tract filter and as such determined by articulator movements. Unfortunately it was not possible to establish clear relationships between formant changes and articulator configurations since the exact properties of the sound source and its behaviour are largely unknown. Therefore future studies will have to pay attention to the precise nature of these relationships and more data on the anatomical as well as physical properties of the parrot vocal apparatus are needed in order to establish a reliable model of sound production in these birds.

In the second half of this thesis I address formant perception by birds in comparison with humans. As shown in chapters 2 and 3 there is convincing evidence that both songbirds and parrots use various articulators to filter the sound produced in the syrinx. Although there are differences in vocal communication between songbirds, parrots and humans the mechanisms of sound production share the principle of active vocal tract filtering, enabling both humans and birds to increase the variety of sounds that can be produced. Following this observation the question arises whether the mechanisms underlying formant perception in particular and frequency modulation in general are also comparable between birds and humans. If so, there is no reason to assume that special mechanisms enabling formant perception evolved in humans as a result of coevolution between speech production and perception. Instead general auditory processing capabilities might be sufficient to allow discrimination of human speech sounds.

Chapter 4 deals with a study investigating speaker normalization in zebra finches using natural human speech obtained from Dutch speaking young adults of both sexes. One of the most remarkable phenomena in human speech concerns our ability to recognize words independent of speaker and strong variation between speakers. Speech scientists have attributed this to the human capacity for intrinsic and extrinsic speaker normalization. Intrinsic speaker normalization accounts for the fact that sounds

which are perceived as the same phoneme can have different acoustic realizations (Liberman *et al.* 1967) by assuming that every speech sample can be categorized using a normalizing transformation (Nearey 1989). At the same time it is well known that there is a speaker effect on speech discrimination initially hampering discrimination across speakers (Creelman 1957; Mullenix *et al.* 1989). This difficulty however is overcome by establishing a reference frame from different speech sound samples (Nearey 1989; Magnus & Nusbaum 2007). We have applied operant conditioning techniques to train zebra finches to discriminate between two naturally produced words, *'wit'* and *'wet'*, that differ mainly in their formant patterns and later transfer this discrimination to unfamiliar voices of (1) the same sex and (2) the opposite sex. All of the eight birds tested were able to discriminate between the words and categorize them independent of speaker identity. Our analysis revealed that the essential clue enabling categorization were the different formant patterns. Furthermore, the birds employed, just like humans, a combination of intrinsic and extrinsic speaker normalization to accomplish the task. This result indicates that the way formants are perceived is either widely spread in the animal kingdom or evolved convergently in birds and humans.

The last chapter of this thesis, **chapter 5**, describes a direct comparison of acoustic cue-weighting in vowel perception in zebra finches and Dutch adults. It has been shown in the past that both Swedish and Canadian-English babies aged three to fifteen months are more sensitive towards low frequency components, i.e. F1, when discriminating vowels (Lacerda 1993, 1994; Curtin *et al.* 2009). This is somewhat surprising since the general notion assumes that the language environment dictates speech perception starting as early as 6 months of age. This might either indicate a universal human bias towards lower frequencies in vowel perception or be the result of maturation of the auditory system. However, it has yet to be explored if these biases are strictly linked to speech sound perception and hence a uniquely human property or a more general characteristic of auditory perception. In a very first attempt to tackle this question we provided both zebra finches and native speakers of Dutch with a Go/NoGo discrimination task using a highly comparable setup. Both groups first had to learn to discriminate between two synthesized words differing only in the embedded vowel sound. The vowels were chosen to differ in F1 as well as F2. In the next step two synthesized 'probe' sounds were added to the discrimination task. Probe sounds were never reinforced and the reactions of the subjects to the probes allowed us to draw conclusions about the way these were perceived. One of the probe sounds had the same F1 frequency as the first stimulus and the same F2 frequency as the second stimulus and vice versa for the other probe

sound. The responses to the probes were strikingly similar in birds and humans and both exhibited a cue-weighting bias towards high frequency components, i.e. F2. This is exactly opposite to what has been found in human babies and firstly demonstrates that cue-weighting is not a uniquely human property tied to speech perception and secondly suggests that a developmental component, likely in the form of auditory maturation, plays an important role in the emergence of such a bias. The major strength of this experiment lies in the use of a highly similar setup for testing birds and humans and therefore makes the results maximally comparable. Furthermore it emphasizes the value of comparative studies across species and ages which should be taken into account when studying mechanisms of speech perception.

Discussion and conclusion

In this thesis I have shown that both zebra finches and monk parakeets use different vocal articulators to modify the sound produced by the syrinx. While in songbirds beak gape and the expansion of the OEC are most important in vocal production, in parakeets tongue movements seem to be the major source of spectral modulation. This is very comparable to human speech production and might be one of the most important parameters of speech imitation by parrots. Based on these observations it can be concluded that sound production mechanisms between birds and humans are more similar than initially assumed. This might suggest convergence in evolutionary patterns. It is conceivable that some of the structures involved in vocal production, such as tongue and beak, initially evolved as part of the food processing system. In that case ecological adaptations for different diets were the driving forces behind the evolution of these structures. At a later point the already existing articulators might have been exploited by the communication systems in order to increase sound variation.

Nevertheless there are remarkable differences in the anatomy and physiology of the sound producing organs as well as in the articulatory patterns. More research based on the current findings could therefore provide detailed models of sound production in both songbirds and parrots.

With regard to speech perception in humans and songbirds I have shown that zebra finches, just like humans, rely on formant patterns to discriminate between highly similar words while at the same time using both intrinsic and extrinsic speaker normalization to categorize words independent of speaker identity. This is an important finding with strong implications for the evolution of formant perception. As has been speculated earlier formant perception likely emerged in a wide range of species serving

to obtain information about an individual's sex, age, size and identity (Ghazanfar *et al.* 2007). Speech might at a later point have exploited this capacity for formant perception eventually leading to a communication system that makes extensive use of formants in order to code linguistic meaning. The finding that human adults and zebra finches exhibit the same cue-weighting bias in vowel perception is in accordance with this hypothesis. Both rely more on F2 frequencies which fall in the most sensitive frequency range in both species when categorizing ambiguous vowels. Human infants on the other hand seem to initially base their discrimination on those frequency components which are spectrally most prominent, namely F1.

In summary I have shown that (song)birds hold the capacity for formant production and perception and that the underlying mechanisms show more similarities between birds and humans than realized before. Both songbirds and parrots can serve as valuable models to address specific questions on the exact nature of these mechanisms and eventually identify selection pressures that might have shaped the evolution of such elaborate vocal communication systems as are only found in humans and birds. Now that some of the mechanisms underlying formant production and perception have been identified, future studies can build on this knowledge to explore more detailed questions concerning e. g. the function of formant patterns in natural birdsong, or the modeling of vocal production and perception mechanisms. Synthesizing songs and manipulating resonance patterns will be important tools to address these questions.

2

Vocal tract articulation in zebra finches

Verena R. Ohms, Peter Ch. Snelderwaard, Carel ten Cate & Gabriël J. L. Beckers

Birdsong and human vocal communication are both complex behaviours which show striking similarities mainly thought to be present in the area of development and learning. Recent studies, however, suggest that there are also parallels in vocal production mechanisms. While it has been long thought that vocal tract filtering, as it occurs in human speech, only plays a minor role in birdsong there is an increasing number of studies indicating the presence of sound filtering mechanisms in bird vocalizations as well. Correlating high-speed X-ray cinematographic imaging of singing zebra finches (*Taeniopygia guttata*) to song structures we identified beak gape and the expansion of the oropharyngeal-esophageal cavity (OEC) as potential articulators. We subsequently manipulated both structures in an experiment in which we played sound through the vocal tract of dead birds. Comparing acoustic input with acoustic output showed that OEC expansion causes an energy shift towards lower frequencies and an amplitude increase whereas a wide beak gape emphasizes frequencies around 5 kilohertz and above. These findings confirm that birds can modulate their song by using vocal tract filtering and demonstrate how OEC and beak gape contribute to this modulation.

Published in PLoS ONE (2010) 5: e11923

Introduction

Birdsong is a complex vocal behaviour often considered to show striking developmental and structural similarities with human speech (Doupe & Kuhl 1999). However, these similarities are mainly thought to be present in the area of development and learning whereas vocal production mechanisms have long been considered to be fundamentally different.

In humans, voiced speech is produced by vibrations of the vocal folds which are subsequently filtered in order to produce different speech sounds that form an important part of our phonetic repertoire (Ladefoged 2006). This filtering process takes place in the upper vocal tract by altering the dimensions of various resonance cavities within the vocal tract, like pharyngeal, oral and nasal cavity. This is achieved by moving articulators such as tongue, lips and lower jaw (Titze 2000)

In contrast to this source-filter theory of human speech (Fant 1960) it has been long thought that frequency and amplitude modulations of bird vocalizations are mainly produced by the avian sound source, the syrinx, and that vocal tract filtering as in human speech production plays a minor role in generating vocal complexity in birdsong (Greenewalt 1968). However, recent studies suggest that this view needs to be reconsidered as there is a growing body of evidence indicating the significance of vocal tract filtering in bird vocal communication as well.

Some of the first evidence derives from experiments showing that both songbirds and non-songbirds singing in heliox exhibit deviating vocal characteristics (Nowicki 1987; Ballentijn & ten Cate 1998). Harmonic overtones of supposedly pure tones become apparent as well as a shifted emphasis towards higher frequencies in broad-band sounds. These observations lead to the hypothesis that the bird's vocal tract can act as an acoustic filter and be actively modulated (Nowicki 1987). Motivated by these findings subsequent studies on potential vocal tract articulators showed that beak movements and gape width are correlated with frequency patterns in white-throated sparrows (*Zonotrichia albicollis*), swamp sparrows (*Melospiza georgiana*) (Westneat *et al.* 1993) Darwin's finches (Podos *et al.* 2004) and zebra finches (*Taeniopygia guttata*) (Williams 2001). Furthermore, in zebra finches correlations between beak gape and amplitude have been found (Williams 2001; Goller *et al.* 2004). Experimentally manipulating beak movements and gape widths also affects frequency patterns in white-throated sparrows, swamp sparrows, canaries (*Serinus canaria*) (Hoese *et al.* 2000) and zebra finches (Goller *et al.* 2004).

Other studies indicate that expanding the oropharyngeal-esophageal cavity (OEC) plays a role in vocal tract filtering as well by tuning it to the fundamental frequency of the vocalizations in doves (*Streptopelia risoria*) (Riede *et al.* 2004), northern cardinals (*Cardinalis cardinalis*) (Riede *et al.* 2006) and white-throated sparrows (Riede & Suthers 2009). Tongue movements in monk parakeets (*Myiopsitta monachus*) also seem to have a filtering effect on the sound produced (Beckers *et al.* 2004).

Although all of the mentioned studies suggest that possible articulators such as beak and the expandable esophagus are likely to modulate birdsong, these data are predominantly correlational. As such, they are insufficient to precisely assess the role of different articulators in vocal production since their effects usually cannot be separated from each other or from other factors such as variation at the sound source. In the current study we combined correlational and experimental data on vocal production in zebra finches. First we used high-speed X-ray cinematographic imaging to quantify patterns of both beak movements and OEC expansion during singing and matched these patterns to distinct note types. Subsequently we conducted an experiment in which we replaced the syrinx by a mini-loudspeaker and played frequency sweeps under varying articulator configurations through the vocal tract (similar to Beckers *et al.* 2004). We manipulated beak gape and OEC expansion and compared acoustic input with acoustic output in order to evaluate the significance of these possible articulators.

Material and Methods

Ethics statement

All animals came from the Leiden University breeding colony and were housed in groups of at least two birds prior to the experiments. All animal procedures were approved by the animal experimentation committee of Leiden University (DEC numbers 08116 and 07190).

Subjects

We used five male and two female zebra finches for the X-ray cinematographic imaging and five male zebra finches for the experiment in which we replaced the syrinx by a mini-loudspeaker. The female birds only served as stimulus birds in the X-ray setting to stimulate the males to sing. During X-ray recordings male birds were individually transferred into a small cage (30 cm wide x 20 cm high x 10 cm deep) built from wood

with plexi glass on both long sides. The small size of the cage allowed to optically focus on the birds, but still allowed the typical dancing movements during singing (Williams 2001).

Cineradiography

A Philips Optimus M 200 X-ray apparatus was combined with a Kodak Motion Corder Analyzer SR- 500 s that records at 500 field s^{-1}, shutter speed 1/500 s by replacing the original camera of the X-ray apparatus by the Kodak system. The images which had a resolution of 512 x 240 pixels were loaded into the camera's onboard memory. The maximum recording time of the Kodak Motion Corder which was triggered manually is 8.7 s at 500 fields s^{-1}, making it necessary to save the video sequences immediately on digital video for permanent storage (Snelderwaard *et al.* 2002). For that we used a Sony Mini Digital Video cassette recorder Model No.GV-D900E and later on an LG DVD Player (DVD Player ± RW Recorder) Model No.DR6621 on which simultaneously sound was recorded too using a pre-amplifier (Marantz PMD661) and a directional microphone (Sennheiser ME 67/ K6) aimed at the bird from 0.5 m distance. As these devices have a frame rate of only 25 frames s^{-1} we played back the video sequences from the Kodak system with 25 frames s^{-1} to prevent data loss while re-recording. We continuously applied an X-ray dose of 56 kV, 60 mA. The videos were captured either from the Mini DV tapes or from the DVDs using Adobe Premiere Pro software version 7.0 for Windows. Due to a distinct tone produced by the X-ray apparatus only while the shutter was open it was possible to align sound and video using the frame-matching features of Adobe Premiere with an accuracy of 2 milliseconds.

In order to accurately follow and quantify the movements of certain articulators we also glued several lead markers (ca. 0.5 mm^3) on head and beak of the birds using tissue adhesive (Superglue 90-120 CPS, World Precision Instruments, Inc., Sarasota, Florida, USA). In two birds we implanted additional lead markers into the tongue and larynx. These procedures were conducted under anesthesia using isoflurane (1.8 %, O_2 0.3 l/min, N_2O 0.4 l/min). To quantify beak gape and OEC expansion we measured the distance between the tips of mandible and maxilla and the distance between the most ventral point of the OEC and the midpoint of the neck of the bird for several note types per song and always at the temporal midpoint of those notes as identified on song spectrograms. These measurements were taken from still images using AviDigitiser (© Peter Ch. Snelderwaard) which provides the coordinates of manually selected points within each video frame and from which distances can be calculated. Only X-ray images

in which the birds kept their head in a perfectly lateral position towards the camera were used, so that in the end we obtained measurements from at least 8 songs for every measured note type. Although we recorded with a high frame rate, the stereotyped dancing movements of zebra finches allowed sampling of only a few notes per song per bird, namely those in which the bird's head was perfectly lateral to the camera. This method is well suited for comparing beak gape and OEC expansion for various elements within a song, as well as for qualitative comparison between birds, but not for quantitative comparisons between birds. Furthermore, due to the small size of zebra finches it was not possible to identify structures such as single vertebrae, but the overall shape of OEC and neck was used to obtain measurements.

Afterwards we carried out a direct discriminant function analysis using beak gape and OEC expansion as predictors for determining which note types were produced. Since zebra finches produce several different and complex note types we did not relate our articulation measurements to signal analytic features such as fundamental or peak frequency as has been done in other studies (Riede *et al.* 2006; Riede & Suthers 2009). Little is known about how zebra finch note types are produced physiologically, but it seems likely that they correspond to different syringeal production modes. Correlating parameters such as fundamental frequency with vocal tract articulation should therefore not be based on an analysis that mixes replicate notes of different types, but rather on one that distinguishes within- and between note type variation. For such an analysis, however, more data would be necessary. At the same time articulatory states of beak gape and OEC expansion might be different enough between various note types to allow predicting which note types relate to different articulator configurations using a discriminant function analysis, although note types exhibiting similar articulatory modes are less likely to be classified correctly.

Prior to the X-ray cinematography we recorded the songs of each bird in a sound-attenuating chamber (ca. 1.80 m x 1.20 m x 2.00 m) that was lined with acoustic foam (Gamma geluidsisolatie platen product number 102247, Intergamma B.V. Leusden, The Netherlands) to reduce sonic reflections from the walls. From these recordings we later took amplitude measurements using the software Praat (version 4.6.09, freely available at www.praat.org) (Boersma 2001) of those note types for which we also measured beak gape and OEC expansion on the X-ray videos. We took care to always take measurements from the temporal midpoint of each note as identified on sound spectrograms in both X-ray videos and song recordings.

However, since X-ray cinematography does not allow evaluating the effects of

beak gape and OEC expansion separately from the sound source the second experiment was conducted to assess a causal relationship and to examine the role of each of these structures in vocal tract filtering directly.

Speaker experiment

One observation made on the X-ray videos is that OEC expansion is caused by a posterior-ventral movement of the hyoid skeleton. Therefore we posterior-ventrally displaced the hyoid skeleton in 0.5 mm steps to gradually increase OEC expansion and evaluate its filtering characteristics while playing frequency sweeps through the vocal tract of freshly sacrificed zebra finches. We did so for three different beak gapes.

The birds used for this experiment were euthanized with an overdose of Nembutal (300 mg/kg body weight) in the pectoral muscle. Afterwards a small incision was made posterior from the lower jaw to expose the urohyal bone (Heidweiller & Zweers 1990) of the tongue apparatus. A cord was knotted around this bone which was later attached to a micromanipulator that could be moved in 0.5 mm steps. Subsequently, the syrinx and a part of the trachea were made accessible by dissecting the birds ventrally between the clavicles following the sternum. The trachea was intersected just above the splitting into the two primary bronchi and a short silastic tube which was fitted over the port of a small speaker (Knowles WBHC NB-68438C, Itasca, Illinois, USA) was inserted into the trachea so that the speaker was placed in the same position where otherwise the syrinx would have been (Beckers *et al.* 2004). The dissected tissue was then agglutinated with tissue adhesive (Superglue 90-120 CPS, World Precision Instruments, Inc., Sarasota, Florida, USA) and the head of the bird was fixed in a stereotaxic device in such a way that the bill was positioned vertically. A thin metal wire (0.7 mm diameter) was stuck between the tips of mandible and maxilla and fixed with tissue adhesive to keep the beak gape constant. During the experiment acoustic measurements with three different beak gapes were taken. In the first series the beak was kept open at ca. 4.0 mm which represented a wide opening as observed on the X-ray videos only during some notes. In the second series the beak was kept open at 1.0 mm, a range frequently observed during natural zebra finch song. In the third series the beak was closed completely. Within each series the position of the hyoid skeleton was changed stepwise by displacing the urohyal bone ventrally in 0.5 mm steps in order to model the expansion of the OEC as observed on the X-ray videos. The maximal ventral movement of the urohyal bone varied between birds and series with a minimal displacement of 4.0 mm and a maximal displacement of 6.5 mm.

The acoustic measurements took place in the sound-attenuating chamber described above. For every position of the tongue apparatus within all three series a linear frequency sweep (0.3 to 10 kHz in 1 second) constructed with PRAAT was played through the vocal tract of the birds using a sound card (CDX-01 CardDeluxe, Digital Audio Labs, 1266 Park Road Chanhassen, MN 55317). The sound emitted from the beak was then recorded with a Sennheiser MKH50 microphone vertically directed at the beak from 3 cm distance and immediately recorded in PRAAT with the same sound card (44.1 kilosamples/s, 16 bit resolution). After the experiment we checked for every bird whether the speaker was still attached to the trachea, which was the case for all five birds. To ensure that differences between spectra of recorded sweeps were caused by differences in articulation and not by position-dependent filtering due to remaining room resonances, we took care that the exact position of both the microphone and the bird preparation did not change between recordings. We also measured speaker output at approximately the same position where the beak was during recordings in order to correct for frequency response deviations of the speaker system by subtracting the dB values of the speaker output from the measured spectrum. Although remaining resonances might still affect the data slightly this impact can be considered rather small and does not change the general results.

The data were analyzed by calculating the long-time average spectrums (Ltas function in Praat; 100 Hz bin width) of the recorded sound sweeps, and comparing them between different articulatory states. The latter was done using custom-written scripts in the scientific computing environment SciPy version 0.7 (Jones *et al.* 2001).

Results

Cineradiography

We obtained sufficient video data from four male birds and measured beak gape and OEC expansion of different note types within each song per individual. Figures 2.1 to 2.4 suggest that different note types are characterized by different combinations of beak gape and OEC expansion.

For every bird separately a direct discriminant function analysis was carried out (Tables 2.1-2.4) with beak gape and OEC expansion as predictors for distinct note types. Two discriminant functions were calculated both of which are significant in birds 498 and 705. In the other two birds, 499 and 704, only the first, but not the second, discriminant functions are significant (Table 2.1).

Table 2.1. Statistical significance of discriminant functions

Bird	Test of function(s)	Wilks' Lambda	Chi-square	df	p
498	1 through 2	0.221	42.224	6	**0.000**
	2	0.517	18.478	2	**0.000**
499	1 through 2	0.542	68.993	12	**0.000**
	2	0.959	4.753	5	0.447
704	1 through 2	0.573	50.736	6	**0.000**
	1	0.962	3.480	2	0.176
705	1 through 2	0.283	91.443	8	**0.000**
	2	0.722	23.611	3	**0.000**

This table gives Wilks' lambda for the two discriminant functions, using beak gape and OEC expansion as parameters, calculated for every bird separately and the chi-square values into which Wilks' lambda can be transformed as well as the corresponding p-values. Significant p-values are printed bold. df, degrees of freedom.

In all four birds beak gape is weighted heavier in the first discriminant function whereas OEC expansion is weighted more in the second, as shown by the standardized coefficients and the correlation between each variable and any discriminant function (Table 2.2).

The classification results (Tables 2.3 and 2.4) show that the percentage of cases in which note types were correctly classified as belonging to their own group is generally well above the percentage expected by chance although some note types were not correctly assigned. Especially in bird 499 (Fig. 2.1, Table 2.3) in which seven different note types were measured the classification for three of these note types remained around chance level, whereas in the other three birds always one note type appeared to be difficult to assign to the right group. In bird 705 note 1 was never properly allocated (Fig. 2.4, Table 2.4) which can be explained by the large overlap between this note and notes 3 and 5. However, the average value of note 1 is closer to the average value of note 3 compared to 5 while at the same time both notes represent harmonic stacks with a comparable sound shape. In bird 499 notes 1 and 3 show a similar structure regarding frequency modulation with the highest amplitude in the lowest frequency band although note 3 has a slightly higher fundamental frequency and a longer duration (Fig. 2.1). At the same time both notes show an almost equal degree of a relatively large OEC expansion and

Table 2.2. Discriminant function coefficients and within-group correlations.

		Coefficients		Correlation	
Bird	**Variable**	**Function 1**	**Function 2**	**Function 1**	**Function 2**
498	Beak gape	0.994	-0.416	1.000	-0.014
	OEC expansion	0.16	1.078	0.386	0.922
499	Beak gape	0.910	0.430	0.850	0.527
	OEC expansion	-0.530	-0.855	-0.427	0.904
704	Beak gape	1.012	-0.088	0.996	0.092
	OEC expansion	-0.093	1.012	0.087	0.996
705	Beak gape	0.974	0.305	0.876	0.482
	OEC expansion	-0.492	0.894	-0.299	0.954

This table lists the standardized canonical discriminant function coefficients and the pooled within-groups correlations between discriminating variables (beak gape and OEC expansion) and both discriminant functions for every individual bird. In all four birds beak gape is the larger standardized coefficient in the first discriminant function and also has the stronger correlation, whereas in the second discriminant function OEC expansion is the larger standardized coefficient and also shows the stronger correlation. OEC, oropharyngeal-esophageal cavity.

only a slight difference in beak gape. These examples therefore fulfill the expectation that similar note types can be characterized by similar articulator configurations.

While it is difficult to compare note types between individuals since the four males sing different note types, it is possible to detect some similarities between birds that produce comparable note types. Bird 498 (Fig. 2.2) for instance produces a frequency modulated note (note 4) with an upwards sweep in the second half of the note which is comparable to note 5 in bird 705 (Fig. 2.4). In both cases these notes show a relatively high OEC expansion and a similar beak gape. Another comparison can be drawn in individuals 499 (Fig. 2.1) and 498 (Fig. 2.2) since they produce a rather dense harmonic stack (note 4 in bird 499, note 3 in bird 498) and in both birds these elements are produced with a relatively large beak gape, although OEC expansion varies remarkably. This again might indicate that the frequency pattern is more influenced by beak gape. On the other hand three of the birds produce high notes (note 1 in bird 498, note 5 in bird 499 and note 2 in bird 705) and in all three cases OEC expansion is at a minimum.

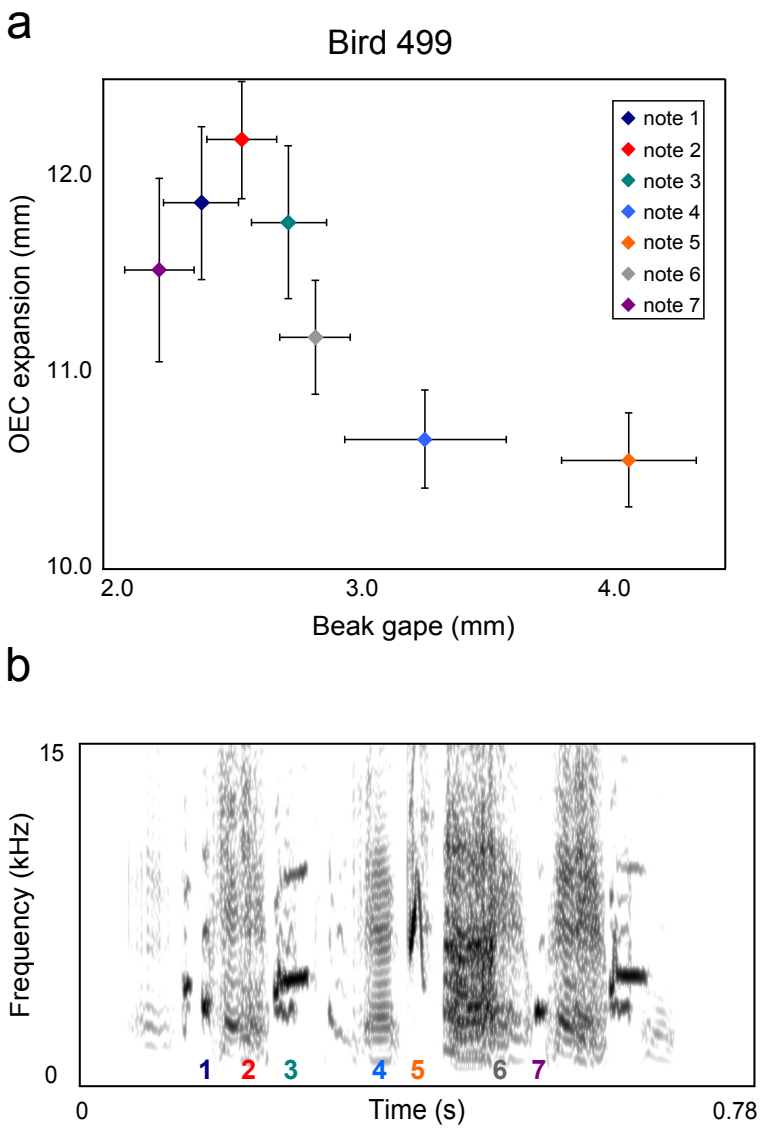

Figure 2.1. Scatter plot of measured note types and song spectrogram for bird 499.

This figure illustrates the results of the measurements taken from the X-ray videos of bird 499. In panel **(a)** average OEC expansion (in millimeters) is plotted against average beak gape (in millimeters) including standard error for several distinct note types measured from at least 8 songs per note. Panel **(b)** shows the associated spectrogram. The numbers below the notes in the spectrograms correspond to the plotted notes in panel (a). mm, millimeters; kHz, kilohertz; s, seconds; OEC, oropharyngeal-esophageal cavity.

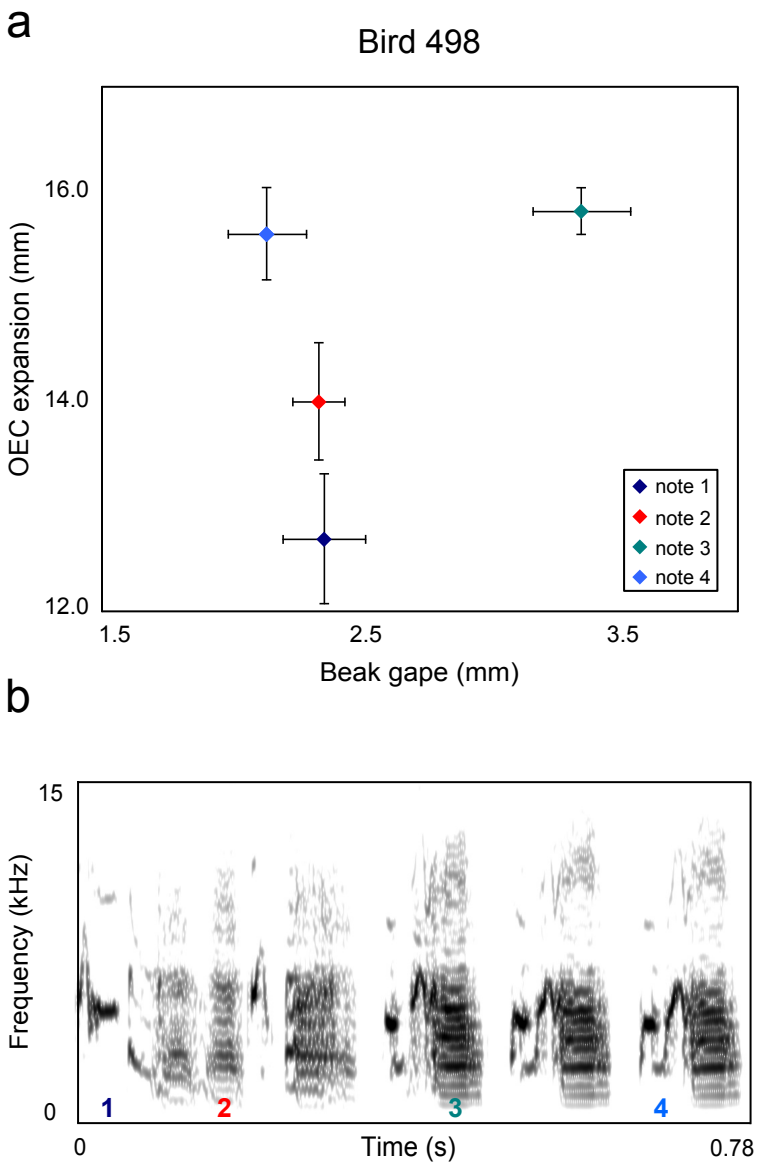

Figure 2.2. Scatter plot of measured note types and song spectrogram for bird 498.

This figure is equivalent to figure 2.1 and shows the results of the measurements taken from the X-ray videos of bird 498. Again panel **(a)** gives a scatter plot of average beak gape versus average OEC expansion including standard error with the associated spectrogram in panel **(b)**. mm, millimeters; kHz, kilohertz; s, seconds; OEC, oropharyngeal-esophageal cavity.

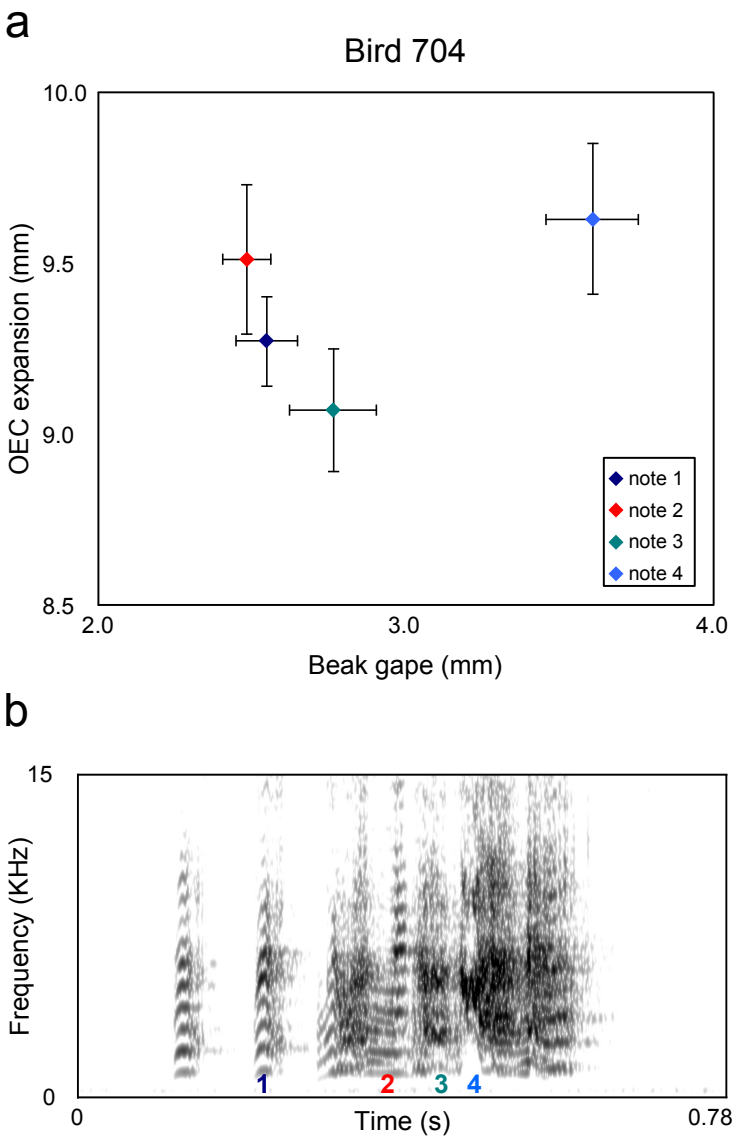

Figure 2.3. Scatter plot of measured note types and song spectrogram for bird 704.

This figure is also equivalent to figure 2.1 and shows the results of the measurements taken from the X-ray videos of bird 704. Again panel **(a)** gives a scatter plot of average beak gape versus average OEC expansion including standard error with the associated spectrogram in panel **(b)**. mm, millimeters; kHz, kilohertz; s, seconds; OEC, oropharyngeal-esophageal cavity.

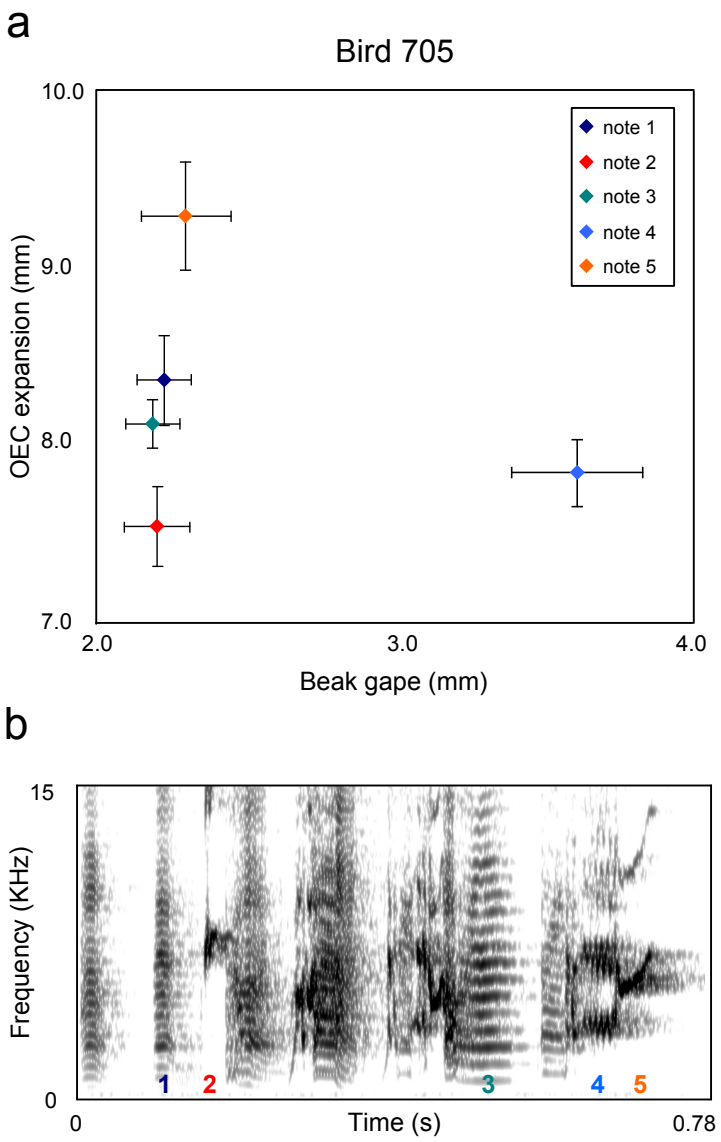

Figure 2.4. Scatter plot of measured note types and song spectrogram for bird 705.

This figure is also equivalent to figure 2.1 and shows the results of the measurements taken from the X-ray videos of bird 705. Again panel **(a)** gives a scatter plot of average beak gape versus average OEC expansion including standard error with the associated spectrogram in panel **(b)**. mm, millimeters; kHz, kilohertz; s, seconds; OEC, oropharyngeal-esophageal cavity.

Table 2.3. Classification results for birds 498 and 499.

Bird 498

	Predicted group membership in % (counts)					
Note	**1**	**2**	**3**	**4**	**N**	**Chance (%)**
1	62.5 (5)	25.0 (2)	12.5 (1)	0.0 (0)	8	25
2	37.5 (3)	25.0 (2)	0.0 (0)	37.5 (3)	8	25
3	0.0 (0)	0.0 (0)	100.0 (8)	0.0 (0)	8	25
4	0.0 (0)	12.5 (1)	0.0 (0)	87.5 (7)	8	25

Bird 499

	Predicted group membership in % (counts)								
Note	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**N**	**Chance (%)**
1	17.6 (3)	41.2 (7)	0.0 (0)	17.6 (3)	0.0 (0)	0.0 (0)	23.5 (4)	17	14.4
2	5.9 (1)	41.2 (7)	11.8 (2)	5.9 (1)	0.0 (0)	11.8 (2)	23.5 (4)	17	14.4
3	5.9 (1)	29.4 (5)	17.6 (3)	11.8 (2)	0.0 (0)	17.6 (3)	17.6 (3)	17	14.4
4	0.0 (0)	6.3 (1)	6.3 (1)	6.3 (1)	31.3 (5)	31.3 (5)	18.8 (3)	16	13.6
5	0.0 (0)	0.0 (0)	5.9 (1)	5.9 (1)	70.6 (12)	11.8 (2)	5.9 (1)	17	14.4
6	5.9 (1)	17.6 (3)	11.8 (2)	11.8 (2)	11.8 (2)	29.4 (5)	11.8 (2)	17	14.4
7	5.9 (1)	29.4 (5)	0.0 (0)	17.6 (3)	0.0 (0)	5.9 (1)	41.2 (7)	17	14,4

In this tabel the percentages as well as the actual numbers of cases in which a note type has been correctly classified as belonging to its own group or misclassified as belonging to another note type are given for birds 498 and 499. In the last column the percentage for a certain note type being correctly identified by chance is listed.

Another factor that might be influenced by beak gape and OEC expansion is amplitude and indeed three out of four birds produce the loudest note measured with the largest OEC expansion and in two cases also with a wide beak gape (Figs 2.1-2.4, Table 2.5).

Generally speaking the results indicate that beak gape as well as OEC expansion might act as vocal tract articulators to generate different note types within each zebra finch song. However, no clear picture concerning the specific effects of these articulators on sound modulation is emerging yet and the speaker experiment was carried out to directly address the role of beak gape and OEC expansion.

Table 2.4. Classification results for birds 704 and 705.

Bird 704

	Predicted group membership in % (counts)					
Note	**1**	**2**	**3**	**4**	**N**	**Chance (%)**
1	40.0 (10)	32.0 (8)	24.0 (6)	4.0 (1)	25	26.3
2	29.2 (7)	58.3 (14)	8.3 (2)	4.2 (1)	24	25.3
3	23.8 (5)	14.3 (3)	28.6 (6)	33.3 (7)	21	22.1
4	0.0 (0)	12.0 (3)	12.0 (3)	76.0 (19)	25	26.3

Bird 705

Note	**1**	**2**	**3**	**4**	**5**	**N**	**Chance (%)**
1	0.0 (0)	23.1 (3)	38.5 (5)	0.0 (0)	38.5 (5)	13	16.9
2	0.0 (0)	60.0 (9)	26.7 (4)	6.7 (1)	6.7 (1)	15	19.5
3	0.0 (0)	17.6 (3)	64.7 (11)	0.0 (0)	17.6 (3)	17	22.1
4	0.0 (0)	12.5 (2)	6.3 (1)	81.3 (13)	0.0 (0)	16	20.8
5	0.0 (0)	12.5 (2)	18.8 (3)	6.3 (1)	62.5 (10)	16	20.8

This table is equivalent to table 2.3 and shows the classification results per note type for birds 704 and 705.

Table 2.5 Amplitude values of measured song elements.

	Bird 498	**Bird 499**	**Bird 704**	**Bird 705**
Note 1	62.64	61.13	58.64	54.93
Note 2	60.95	56.20	59.45	56.04
Note 3	74.09	65.76	64.70	62.69
Note 4	66.27	55.72	73.11	68.26
Note 5		61.87		69.05
Note 6		64.22		
Note 7		62.02		

Table 2.5 gives the amplitude values in decibel of all song elements for which beak gape and OEC expansion have been measured based on the X-ray videos.

Speaker experiment

The results of this experiment are displayed in figure 2.5 which shows the effect of varying beak gape and OEC expansion on vocal tract resonances in five individual zebra finches independent of the acoustic characteristics of the syringeal sound source.

The overall frequency and amplitude modulation pattern is rather similar in all five individuals indicating a fairly specific effect of the mentioned articulators on the sound signal. The clearest and most consistent effect is given by the expansion of the OEC leading to an amplitude increase especially in the frequency region around 5 kHz which is particularly obvious at a wide beak gape, but also clearly visible when the beak is only slightly open. The total amplitude increases on average 10.9 dB when the beak is only slightly open and 8 dB when the beak is wide open. With a closed beak this effect is not so clear since amplitude first increases with OEC expansion but at a certain point drops again; in three birds even below the intensity level that was gained at position 0. Also at the other two beak gapes amplitude first increases rapidly, but drops again at a certain point although it clearly remains above the intensity level at the beginning of a series. Furthermore OEC expansion causes an energy shift towards relatively lower frequencies which again is especially obvious with a wide beak gape, but also visible in the two other conditions.

Beak gape has a strong effect on amplitude too. The wider the beak opens the louder the sound becomes with an average amplitude increase of 5.8 dB compared between a closed and a wide beak gape.

Regarding frequency range it seems that a closed beak filters frequencies above 6 kHz whereas a wide beak gape emphasizes high frequencies above 5 kHz and over a broader frequency range.

All of the nine subplots show a low prominent formant around 1.5 kHz probably mostly influenced by the resonating trachea which dimensions do not change under the different articulator configurations tested in this experiment.

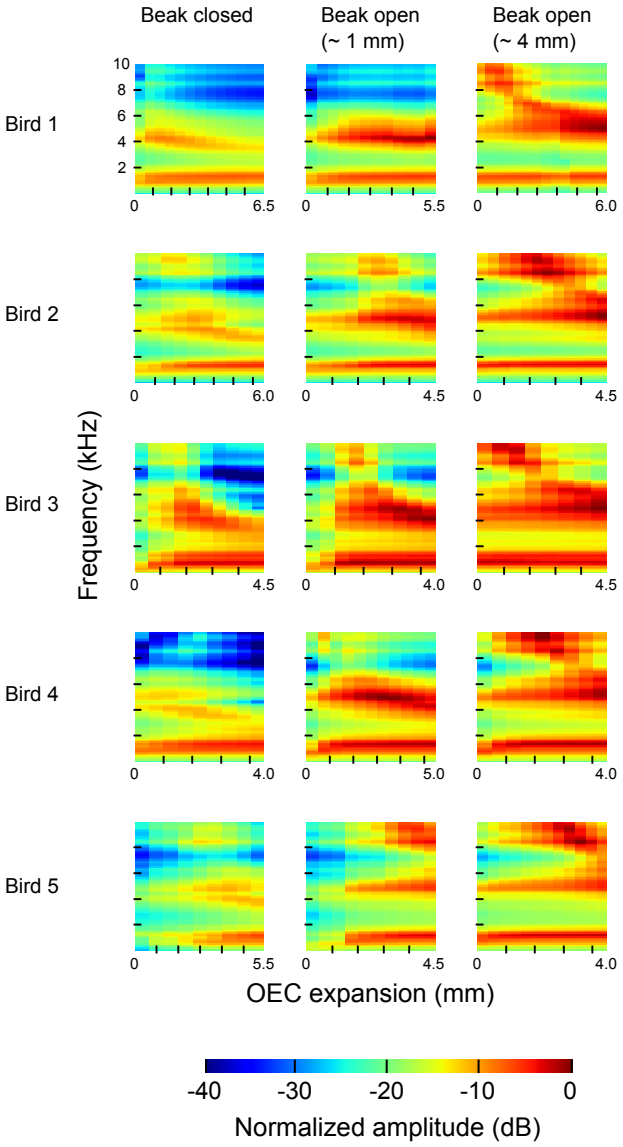

Figure 2.5 Resonance characteristics of the zebra finch vocal tract as a function of beak gape and OEC expansion.

Each subplot represents the sound energy density as a function of sound frequency and OEC expansion, which means that within a single subplot the effect of OEC expansion on the sound is illustrated. Three subplots in a row belong to one individual bird and differ from each other by the degree of beak opening. Red-orange areas represent frequencies with relatively high sound levels and therefore correspond to vocal tract resonances. This figure has been corrected for deviations in the frequency response of the speaker system. mm, millimeters; kHz, kilohertz; dB, decibel; OEC, oropharyngeal-esophageal cavity.

Discussion

In the current study we combined observational and correlational data on song production in zebra finches with findings derived from experimental manipulations of beak gape and OEC expansion to provide insight into the mechanisms of vocal tract filtering in this species. Based on our results it seems clear that zebra finches can use both beak gape and OEC expansion as vocal articulators to filter the sound produced by the syrinx. However, while some of our results support conclusions drawn by other studies and are in line with some of the hypotheses formulated earlier, not all our findings confirm what has been discovered regarding vocal tract filtering in other bird species.

Our X-ray cinematographic imaging of singing zebra finches revealed that the expansion of the OEC is caused by a cyclical posterior-ventral movement of the hyoid skeleton which is comparable to northern cardinals (Riede *et al.* 2006) and white-throated sparrows (Riede & Suthers 2009) which both increase the volume of their oropharyngeal-esophageal cavity by cyclically moving the hyoid skeleton. In these species the OEC functions as a resonance cavity that tracks the fundamental frequency of the song which is in accordance with our data from the speaker experiment showing a downwards-shift in peak frequency with increasing OEC expansion (Fig. 2.5). However, it seems that in zebra finches OEC expansion also causes a general amplitude increase independent of specific frequencies which is especially obvious at a wide beak gape (Fig. 2.5). This does not only become apparent in the speaker experiment but also gains support by the X-ray videos since three of the birds produce the loudest note measured with the largest OEC expansion (Figs 2.1-2.4, Table 2.5). In this context it is also interesting to note that reduced air sac volume in zebra finches causes sound amplitude to decrease whereas the temporal pattern of the song remains unaffected (Plummer & Goller 2008).

Beak gape has been shown to essentially influence frequency patterns of bird vocalizations although different studies arrive at different conclusions. The general picture emerging from the literature is that wider beak gapes correlate with higher frequencies whereas smaller beak gapes correlate with lower frequencies. This has been shown in several songbird species such as the white-throated sparrow, the swamp sparrow (Westneat *et al.* 1993) and the song sparrow (*Melospiza melodia*) (Podos *et al.* 1995) but also in barnacle goose (*Branta leucopsis*) (Hausberger *et al.* 1991) Support for a causal relationship of these observations comes from experiments in which beak gape was experimentally manipulated (Hoese *et al.* 2000) either by immobilizing or by adding weight to the beak. In the latter case lower frequency notes were more strongly affected

which in turn is consistent with findings from theoretical and experimental modeling (Fletcher & Tarnopolsky 1999) that predict a non-linear relationship between beak gape and vocal tract resonances in a way that changes at smaller beak gapes lead to relatively larger changes in vocal tract resonances.

Another study (Nelson *et al.* 2005) confirmed in eastern towhees (*Pipilo erythrophthalmus*) the hypothesis that beak gape articulation causes significant modulation of the vocal tract filtering function. In this species frequencies between 4 and 7.5 kHz are attenuated when beak gape width is reduced. Furthermore the authors propose that towhees in particular and songbirds in general might vary beak gape as a mechanism to exclude or concentrate energy in distinct frequency bands which often results in the production of narrow-band or pure-tone sounds.

Based on the results of our speaker experiment it seems that on the one hand large beak gapes indeed sustain high frequencies (Fig. 2.5) and thereby partly confirm what other studies have found (Westneat *et al.* 1993; Podos *et al.* 1995) while at the same time a closed beak attenuates frequencies above 6 kHz. However, the analysis of the X-ray videos provides ambiguous results since high-frequency notes are not always produced with a large beak gape (Figs 2.1-2.4).

Another observation made in song sparrows is that coordinated beak movements develop rather late during song learning and appear to correspond with improved tonal quality of the sound produced whereas they are not necessary for producing the acoustic fine structure of notes (Podos *et al.* 1995). However, zebra finches mostly produce complex notes with energy distributed over a large range of frequencies with the fundamental frequency often being attenuated, instead of pure-tone sounds while rapid beak movements occur during the whole song. Therefore it seems unlikely that this species adjusts beak gape to improve tonal quality. In fact, Williams (2001) reports a high increase in peak frequency (~ 694 Hz) after beak opening movements whereas the fundamental frequency was only slightly increased (~ 12 Hz). Also the average amplitude was greater after beak opening movements. Our results corroborate these findings. The speaker experiment revealed an amplitude increase with both OEC expansion as well as beak opening. At the same time peak frequency is higher when the beak is open compared to when it is closed. However we could not detect a clear effect of beak gape on peak frequency based on the X-ray data. On the one hand this might be attributed to the fact that other parameters, such as syringeal muscle activity could play a role in frequency modulation too, while on the other hand the sampling rate might be too low since only some notes could be measured per bird.

A different study found a strong positive correlation between beak gape and fundamental frequency as well as peak frequency in zebra finches in most of the individuals tested although the relationship between beak gape and fundamental frequency did not apply to harmonic stacks (Goller *et al*. 2004). The authors also found a correlation between beak gape and amplitude although they conclude from their data that this relationship is likely secondary and based on a strong correlation between air sac pressure and beak gape (Goller *et al.* 2004). Our X-ray data confirm that those notes produced with the largest beak gape usually have a high amplitude (Figs 2.1-2.4, Table 2.5) while the speaker experiment too indicates that beak gape has a strong effect on amplitude and therefore cannot be regarded secondary.

Other structures that might be involved in vocal tract filtering include the trachea itself and glottal opening. Whereas the role of glottal opening has not been examined yet there are indications that zebra finches actively shorten the trachea at the beginning of a song bout (Daley & Goller 2004). However, modulation of tracheal length during the song motif seems to be driven by air sac pressure changes and does not clearly relate to the acoustic structure of the song. Some passerine species such as the trumpet bird (*Phonygammus keraudrenii*) exhibit elongated tracheas which are assumed to lower the pitch of the vocalizations (Clench 1978) and therefore exaggerate size (Fitch 1999).

Figure 2.5 shows in every subpanel a low formant which exhibits basically no frequency modulation and is likely mostly influenced by the trachea of the birds which dimensions do not change during the experiment and therefore remains resonating at a certain frequency. However, this might be different during real vocalizations although the study mentioned above (Daley & Goller 2004) did not find a clear relationship between tracheal length and song structure.

In any case it has become clear that vocal tract filtering in birds can enhance vocal complexity and serve to code biologically relevant information such as size (Fitch 1999; Fitch & Kelley 2000). While it has been thought originally that vocal tract filtering does not apply to birdsong it is obvious nowadays that the source-filter theory of speech production can also be implemented on bird vocal communication. However, given the anatomical and physiological characteristics of the avian sound source we have to assume that the mechanisms underlying vocal production in birds are more complex than those underlying human speech production. On the one hand there is evidence that each side of the syrinx can be controlled independently in at least some songbird species (Suthers 1990; Suthers 1999; Zollinger & Suthers 2008) resulting in e.g. two-voice phenomena. On the other hand it has also been shown that the two syringeal halves

may be coupled and interact with each other (Nowicki & Capranica 1986). Moreover, a multiplicity of syringeal and respiratory muscles controlling airflow and air sac pressure play an important role in generating certain acoustic properties (Goller & Cooper 2004).

In summary we have shown that zebra finches can use both beak gape and OEC expansion to modulate their vocalizations to a substantial degree. However, the wide variety of different note types that these birds produce does not seem to be solely based on the interaction of these articulators but is likely to be affected also by other factors related to the sound source.

Acknowledgements

We thank Arthur J. Schotgerrits for technical support with the video analysis and Thomas Pöhler from Interton Hörgeräte GmbH, Bergisch Gladbach, Germany and Uwe Markert from Audifon GmbH Kölleda, Germany for providing the mini-loudspeakers. Funding was provided by the Netherlands Organization for Scientific Research (NWO). Grant Number 815.02.011.

3

Vocal tract articulation revisited: the case of the monk parakeet

Verena R. Ohms, Gabriël J. L. Beckers, Carel ten Cate & Roderick A. Suthers

Birdsong and human speech share many parallels with respect to vocal learning and development. However, vocal production mechanisms have long been considered to be different. The vocal organ of songbirds is more complex than the human larynx, leading to the hypothesis that vocal variation in birdsong originates mainly at the sound source while in humans is primarily due to vocal tract filtering. However, several recent studies have indicated the importance of vocal tract articulators such as beak and oropharyngeal-esophageal cavity. In contrast to most other bird groups, parrots have a prominent tongue raising the possibility that tongue movements may be of significant importance in vocal production in parrots, but evidence is rare and observations often anecdotal. In the current study we used X-ray cinematographic imaging of naturally vocalizing monk parakeets (*Myiopsitta monachus*) to assess which articulators are possibly involved in vocal tract filtering in this species. We observed prominent tongue height changes, beak opening movements and tracheal length changes suggesting an important role of tongue and beak in producing a resonance cavity and identifying the trachea as another vocal articulator. We also found strong positive correlations between beak opening and amplitude as well as changes in tongue height and amplitude in several types of vocalizations. Our results suggest considerable differences between parrot and songbird vocal production while at the same time parrots vocal articulation might more closely resemble human speech production.

Manuscript

Introduction

In recent years birdsong has become the focus of many scientists interested in the cognitive, neural, genetic and physiological mechanisms underlying human speech and language. The fact that songbirds and humans exhibit many parallels in vocal learning and perception (e.g. Doupe & Kuhl 1999; Ohms *et al.* 2010a) has established songbirds as an excellent model system in which to study the underlying mechanisms in both birds and humans (Bolhuis *et al.* 2010). Also, cognitive mechanisms related to syntax detection might be comparable in humans and songbirds although results are controversial (Gentner *et al.* 2006; van Heijningen *et al.* 2009).

However, while there are numerous analogies there are differences too, especially regarding vocal production. In humans the primary sound source is located in the larynx and voiced speech sounds are produced by a pair of vibrating vocal folds (Titze 2000). The generated acoustic signal is subsequently filtered by shaping the vocal tract using different articulators such as tongue and lips (Ladefoged 2006). This leads to amplification of different frequency regions within the broad-band spectrum of human speech sounds.

The vocal organ of birds on the other hand, the syrinx, is located at the basis of the trachea in the interclavicular air sac (Suthers & Zollinger 2004) and in the case of Oscine songbirds consists of two sets of vibrating labia located at the cranial end of each of the primary bronchi (Goller & Larsen 1997) which are capable of independent motor control (Suthers 1990). This enables songbirds to sing with two voices simultaneously or switch between both sets of labia while singing, depending on the frequencies produced (Suthers 1990; Suthers *et al.* 1994; Suthers *et al.* 2004; Zollinger & Suthers 2004). The more complex vocal organ of songbirds initially led to the hypothesis that acoustic variation predominantly arises at the sound source and that in contrast to human speech acoustic filtering by the vocal tract only plays a minor role in birdsong production (Greenewalt 1968).

Most bird species studied produce relatively narrow-band, tonal songs which lack the complex resonance patterns prominent in human speech. It has been shown, however, that the sound generated at the source can exhibit harmonic overtones (Beckers *et al.* 2003) and that cyclical movements of the hyoid skeleton or expansion of the cervical esophagus filter these out of the signal by tuning the oropharyngeal-esophageal cavity (OEC) to the fundamental frequency of the song (Riede *et al.* 2004; Riede *et al.* 2006; Riede & Suthers 2009). Additionally, in zebra finches (*Taeniopygia guttata*) which

produce a wide range of broad-band note types, expansion of the OEC has also been found to affect frequency patterns by shifting energy to relatively lower frequencies while amplitude generally increases (Ohms *et al.* 2010b). Other articulators involved in avian vocal tract filtering include beak movements and gape widths (Hoese *et al.* 2000; Podos *et al.* 2004; Nelson *et al.* 2005) making clear that there is increasing evidence for the importance of vocal tract filtering in the production of avian vocalizations.

Interestingly, observations of naturally vocalizing and speech-imitating parrots, which have a simpler syrinx with only one pair of vibrating labia (Larsen & Goller 2002) suggest that tongue movements play an important role in vocal production too (Nottebohm 1976; Patterson & Pepperberg 1994; Beckers *et al.* 2004). Compared to songbirds the parrot tongue is morphologically very different in that it contains many intrinsic muscles and its surface is more like the human tongue: a fleshy, rather flexible structure (Homberger 1986) that might be moved in a horizontal and vertical plane within the oral cavity. So far, however, evidence on this subject is rare and observations are often anecdotal. Studies on a speech-imitating African grey parrot *(Psittacus erithacus)* have suggested that this bird can, similarly to humans, adjust the front-back position of its tongue in order to imitate human articulatory patterns while it lacks, contrary to humans, extensive control over the high-low dimension (Patterson & Pepperberg 1994; Warren *et al.* 1996). Another experimental approach evaluating the significance of tongue movements in monk parakeet (*Myiopsitta monachus*) vocalizations has demonstrated that moving the tongue horizontally in the mouth cavity can lead to frequency and amplitude changes in acoustic resonance patterns (Beckers *et al.* 2004). However, no direct observations of tongue movements in naturally vocalizing parrots exist to date nor is it known whether parrots, like songbirds, exhibit a cyclical movement of the hyoid skeleton causing an expansion of the OEC.

In the current study we address these questions by using X-ray cinematographic imaging of the vocal tract during natural vocalizations of monk parakeets. We report on tongue height changes and beak movements during sound production and how these strongly correlate with amplitude. Furthermore we found evidence for tracheal shortening during vocalizing.

Material and Methods

Subjects

The monk parakeets used in this study had been obtained from a U.S. Department of Agriculture pest control program in Florida and were housed in pairs or individually in metal cages (43 cm deep x 44.5 cm wide x 50 cm tall) in the same room under a 14L:10D schedule prior to the experiment. During the experiment all birds were moved in their home cages into the room that contained the X-ray apparatus to stimulate the respective focal bird to vocalize. Food and water were provided *ad libitum* at all times and wooden toys in the cages served as enrichment. X-ray recordings were obtained from four monk parakeets of which three fulfilled our criteria for good lateral views and were included in further analysis.

X-ray cinematography and song recordings

A Series 9800 mobile C-arm and 1 k x 1 k neurovascular work station (OEC Medical Systems, Inc.) was used to obtain X-ray videos of spontaneously vocalizing monk parakeets. This apparatus generated a digital signal of 30 pulses/s and a 1000 x 1000 image resolution. The duration of each X-ray pulse was 10 ms. The focal bird was transferred into a metal cage of the same dimensions given above in which two opposite sides of the cage were replaced by plexiglass panels and enabled recording the bird in a lateral view with the head of the bird being about 5 cm in front of the intensifier screen. The digital signal of the X-ray apparatus was recorded on a Sony GVD-1000 NTSC digital video cassette recorder, mini DV format. Sound was simultaneously recorded using a directional microphone (Audio Technica model AT835b) which was positioned about 0.5 m from the bird. Afterwards relevant sequences of the X-ray movies were digitized and rendered at 30 frames/s (video) and concurrent vocalizations were digitized at 48 kHz sampling rate using the software Vegas Video, Sonic Foundry, Madison, WI, USA, version 5.0. All data files were corrected for a recording delay of approximately 114 milliseconds in the video relative to the audio.

Marker implantation

In all four birds a stainless steel ball (SIS Type 316, 1.59 mm diameter, Small Parts Inc.) with a diameter of 1.59 mm was inserted dorsally under the skin of the neck. This sphere provided a size reference when measuring anatomical distances from the X-ray videos. Additionally, two of the monk parakeets were anesthetized and the trachea was exposed

through a small mid-ventral incision in the skin of the neck and two pieces of silver wire (ca. 2 mm long x 0.16 mm diameter) (Engelhard Fine Wire) were attached with tissue adhesive (3 M Vetbond) to two tracheal rings. These markers were about 13 mm apart in bird 2 and about 10 mm in bird 3. In order to better follow tongue movements during X-ray recordings, we implanted a short piece (ca. 1.5 mm) of the same silver wire into the tongue bottom about 1.5 mm from the tip of the tongue of bird 1. The wire was inserted into the hole made by a 26 ga hypodermic needle and the incision was sealed with a micro-drop of tissue adhesive. All of the described procedures were performed under isoflurane anesthesia administered with a calibrated anesthetic gas vaporizer (Fluotec) through a mask at a concentration of ~1.5 to 2.0% in air.

Anatomical measurements

Only those video sequences in which the birds' heads were clearly laterally oriented towards the X-ray beam were used for measuring anatomical distances during sound production. The distances measured were: (1) 'beak movement' represented by the distance between the dorsal point of the beak- skull transition and the ventral point of the lower mandible where the bones form a strong symphysis, (2) 'tongue height' which was defined as the distance between the tongue's ventral surface measured about 1.5 mm from the tip of the tongue and the same point of the lower mandible as measured in 'beak movement' and (3) 'tracheal shortening' which was determined by changes in the distance between the tracheal markers (Fig. 3.1). These measurements were performed using MaxTRAQ Lite+, version 2.2.0.1 (Innovision Systems Inc.) by manually selecting points of interests in each successive frame. From the coordinates of each selected point distances were automatically calculated between the points. Ten repeated measures of beak movement in the same frame had a standard deviation of 0.12 mm whereas the distance measured between two metal bars had a standard deviation of 0.14 mm.

Acoustic measurements were done with sound analysis software (Praat, version 4.6.09, freely available at www.praat.org; Boersma 2001).

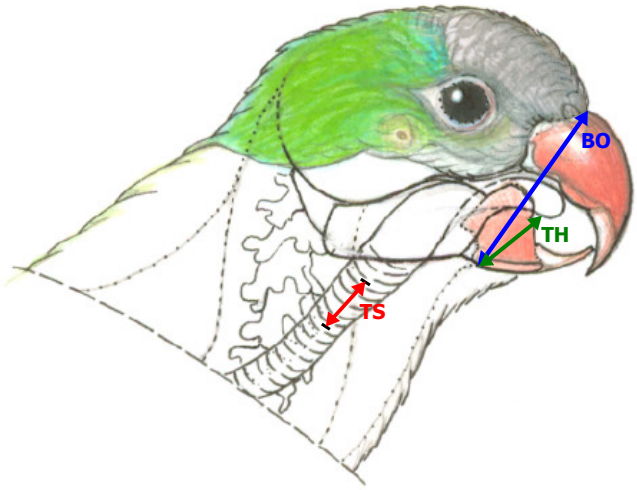

Figure 3.1. Anatomical measurements.

Lateral view of a monk parakeet indicating the distances measured. Beak opening (BO) describes the distance from the dorsal edge of the beak-skull transition to the ventral edge of the lower mandible where the bones form a strong symphysis. Tongue height (TH) is defined by the distance between the ventral surface of the tongue about 1.5 mm from the tip and the lower mandible and tracheal shortening (TS) measures the distance between two tracheal markers.

Results

Vocalizations

Adult monk parakeets produce nine different call types in various contexts, e.g. territorial defense, pair bonding and flock integration, which differ in temporal as well as spectral parameters (Martella & Bucher 1990). In the current study, however, only a subset of these vocalizations was uttered during recording sessions, consisting of contact and greeting calls as well as chatter sounds.

The most common call type produced by the monk parakeets in our study was the contact call (Fig. 3.2a), a short (180.66 ms ± 9.17 s.d. between animals), strongly frequency-modulated (FM) call with discrete, harmonically related frequency bands, which is uttered in many contexts by both sexes (Martella & Bucher 1990). We recorded several instances of contact calls of three birds that met the criteria specified in the methods section to be included in the analysis.

The second-most common call produced by the monk parakeets in this study was the greeting call (Fig. 3.3a) which is considerably longer and more variable in duration (455.70 ms ± 234.39 s.d. between individuals) and does not exhibit the fast FM typical for contact calls. It consists of a spectrally complex pattern with amplified frequency bands that are indicative of formants (Beckers *et al.* 2004) and that exhibit some FM, especially at the beginning of a call.

Furthermore, each of the parakeets produced several sounds which are referred to as chatter (Martella & Bucher 1990). These sounds are mostly characterized by short harmonic stacks which at times exhibit some FM. In the case of bird 1 these short harmonic sounds alternate with notes that exhibit fast FM (Fig. 3.4a).

Articulatory movements

All monk parakeets in this study generally showed the same articulatory movements of beak and tongue when producing contact and greeting calls. Although these call types differed from each other in acoustic structure, no obvious difference in the movement patterns of tongue and beak was detected that could explain the acoustic variation and FM between call types.

Beak opening increased substantially before the onset of a contact call and the tongue which usually rests high in the oral cavity, so that it touches the upper mandible, moves downwards and retracts a bit thereby creating a large oral resonance cavity (Fig. 3.5). Just after call onset both beak gape and tongue height reached their maximum mean displacement with beak movement ranging from 5.57 to 6.68 mm and tongue height ranging from 2.92 to 4.31 mm (Table 3.1). This position was maintained for the duration of the call, after which both articulators returned to their original position.

The movement patterns for beak and tongue during greeting calls were rather similar to those described in contact calls. However, in longer greeting calls the initial beak opening movement proceeded more gradually compared to contact calls, reaching its maximal displacement towards the end of the call, while tongue height decreased faster at the beginning of the greeting call and remained low throughout its duration (Fig. 3.3b). Additionally, beak gape did not increase as much as it did during contact calls with a mean maximum displacement ranging from 4.78 to 6.28 mm whereas tongue depression seemed to be slightly higher in two of the birds (Table 3.2). Furthermore it was noticeable that greeting calls were produced over a wide range of intensities and there was a strong relationship between acoustic power and magnitude of articulatory movements (Fig. 3.6 b,e; see below). Therefore we divided greeting calls into two groups

depending on mean acoustic power measured over the whole call. All greeting calls below 66 dB were referred to as 'soft greeting calls' whereas everything above this threshold was simply referred to as 'greeting calls'. Mean maximum beak movement was on average 3.3 times as high in greeting calls compared to soft greeting calls whereas tongue displacement differed on average only by a factor of 1.9 (Tables 3.2 and 3.3). This suggests that tongue height might be relatively more important than beak gape in generating spectral features which are similar in loud and soft greeting calls while beak gape might mainly affect amplitude.

Figure 3.4 shows the cyclical movements of beak and tongue during the production of two alternating chatter sounds. It is apparent that the magnitude of change of both beak opening and tongue height was less in the second and fourth note compared to the first and third. Examining the corresponding video revealed a strikingly opposite pattern of cyclical tongue movement between these two note types. During the production of the first and third note the tip of the tongue and antero-dorsal part of the tongue body first moved caudally following the movement of the lower mandible while the postero-dorsal part of the tongue body remained higher on a vertical axis. However, during the second part of the sound, which consisted of upward FM sweeps, the postero-dorsal part of the tongue body pushed downwards now forming a horizontal plane with the rest of the structure before the anterior part of the tongue moved rostrally to its resting position high up in the mouth cavity touching the upper mandible. In the second and fourth note this pattern was reversed with the postero-dorsal part of the tongue body moving caudally just before the onset of the note. During the first part of the note the rest of the tongue then completed its caudal movement and again formed a horizontal plane with the postero-dorsal part of the tongue body which was lifted a bit during the second part of the note before the tongue as a whole moved rostrally to its resting position.

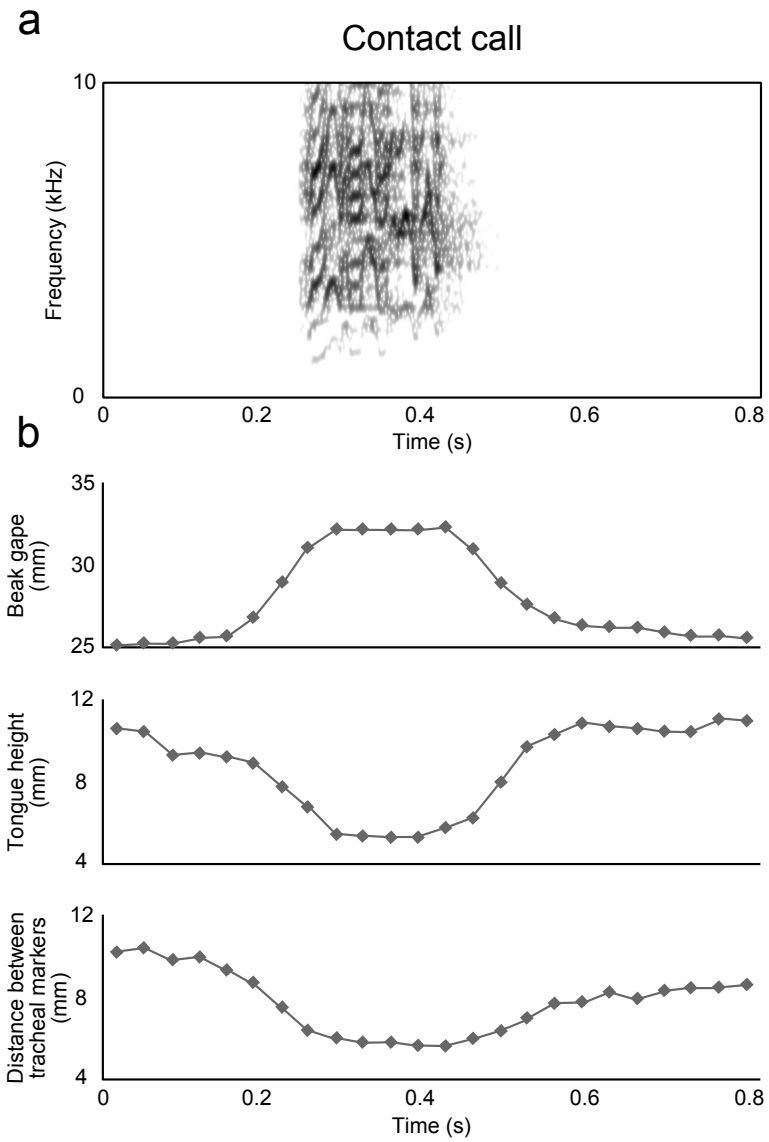

Figure 3.2. Articulatory patterns during contact call production.

Contact calls are accompanied by movements of different articulators. **(a)** Spectrogram of a contact call produced by bird 3. **(b)** Beak opening, tongue depression and tracheal shortening over the course of the contact call displayed in panel (a). kHz, kilohertz; mm, millimeters; s, seconds.

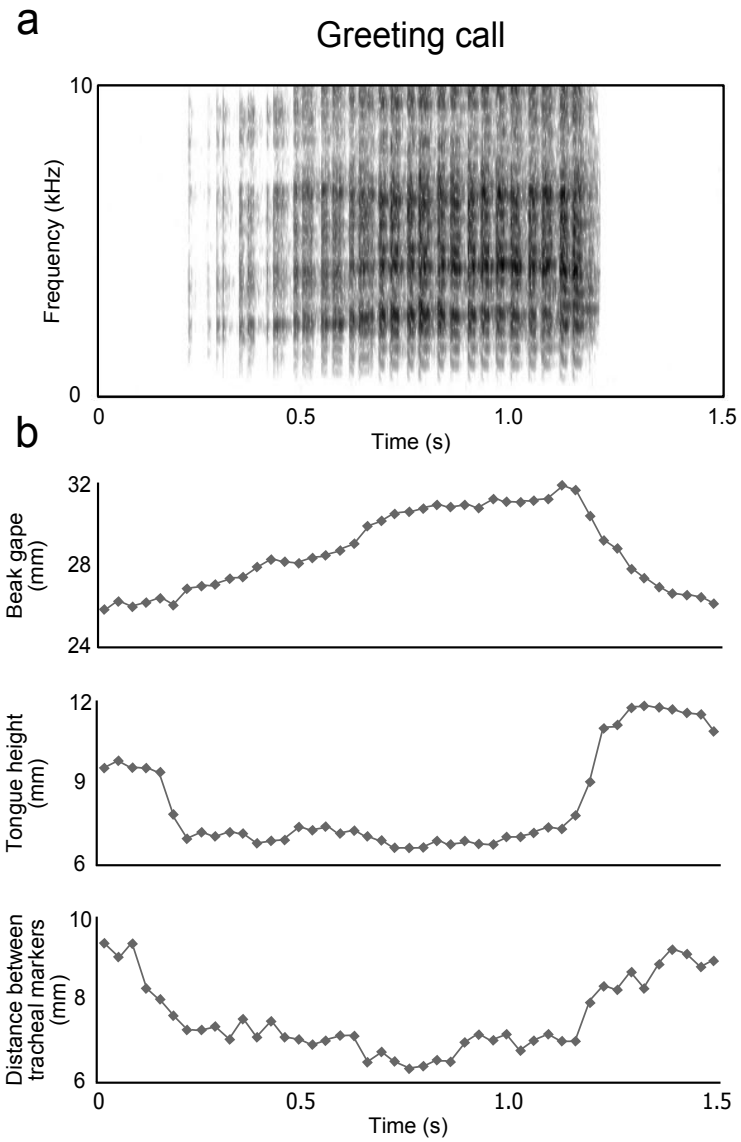

Figure 3.3. Articulatory patterns during greeting call production.

Similar to contact calls, greeting calls are accompanied by changes in the same articulators as described for contact calls. **(a)** Spectrogram of a greeting call produced by bird 3. **(b)** Beak opening, tongue depression and tracheal shortening over the course of the greeting call displayed in panel (a). kHz, kilohertz; mm, millimeters; s, seconds.

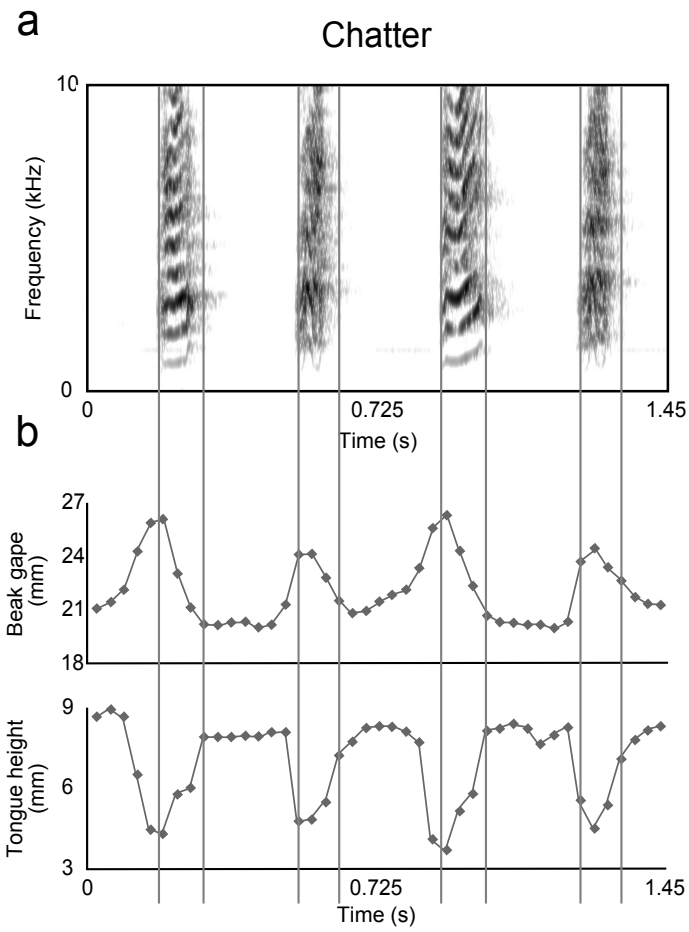

Figure 3.4. Articulatory patterns during chatter sounds.

This figure represents articulatory movements during the production of two alternating chatter sounds of bird 1. **(a)** Spectrogram of chatter sounds. **(b)** Beak opening and tongue depression during the production of the chatter sounds illustrated in panel (a). Note that both beak and tongue reach their maximum displacement just after the onset of the sound while most of the sound is produced when the articulators already move back to their original position. kHz, kilohertz; mm, millimeters; s, seconds.

Changes in tracheal length

In birds 2 and 3 we implanted silver wire markers onto the trachea which could be traced during sound production. In bird 2 these markers were attached to the trachea 18 and 31 mm from the glottis. In bird 3 the markers were implanted 24 and 34 mm from the larynx. The total length of the trachea from glottis to syrinx was 55 mm in bird 2 and 65 mm in bird 3. In both contact and greeting calls the distance between these markers changed substantially over the course of call production with a mean maximum shortening ranging from 5.70 mm in bird 2 to 3.40 mm in bird 3 during contact calls, from 4.78 mm in bird 2 to 3.03 mm in bird 3 during greeting calls and from 1.15 mm in bird 2 to 2.22 mm in bird 3 during soft greeting calls (Tables 3.1-3.3). Postmortem investigation of the trachea revealed that it had very little resistance to substantial changes in length in both birds. Calculating predicted resonances of the tracheas modeled as stopped tubes yields resonances at 1570 Hertz and 1330 Hertz respectively for bird 2 and 3. Both of these values fall within the range of spectral peaks measured over the course of greeting calls.

Table 3.1. Articulator displacement during contact calls.

Bird ID	Beak opening (mm)	Tongue depression (mm)	Tracheal shortening (mm)
1	5.57 ± 0.98; *n* = 6	2.92 ± 0.82; *n* = 6	
2	6.50 ± 1.04; *n* = 10	4.00 ± 0.47; *n* = 10	5.70 ± 0.08; *n* = 2
3	6.68 ± 0.89; *n* = 28	4.31 ± 0.60; *n* = 28	3.40 ± 1.20; *n* = 7

This table lists the mean maximum beak opening movement, tongue depression and tracheal shortening in millimeters per bird inlcuding standard deviation and total number of calls measured. ID, identity; mm, millimeters.

Table 3.2. Articulator displacement during greeting calls.

Bird ID	Beak opening (mm)	Tongue depression (mm)	Tracheal shortening (mm)
1	4.78 ± 1.11; *n* = 4	3.85 ± 1.32; *n* = 4	
2	6.28 ± 0.50; *n* = 4	4.79 ± 1.12; *n* = 4	4.78; *n* = 1
3	5.19 ± 1.15; *n* = 1	4.11 ± 0.92; *n* = 13	3.03 ± 1.24; *n* = 13

This table is equivalent to table 3.1 but lists measurements for greeting calls instead. ID, identity; mm, millimeters.

Table 3.3. Articulator displacement during soft greeting calls.

Bird ID	Beak opening (mm)	Tongue depression (mm)	Tracheal shortening (mm)
2	0.77 ± 0.63; *n* = 6	1.66 ± 0.99; *n* = 6	1.15 ± 0.86; *n* = 2
3	2.70 ± 1.87; *n* = 10	3.03 ± 1.39; *n* = 10	2.22 ± 1.05; *n* = 10

This table is equivalent to tables 3.1 and 3.2 and gives articulator measurements for soft greeting calls. ID, identity; mm, millimeters.

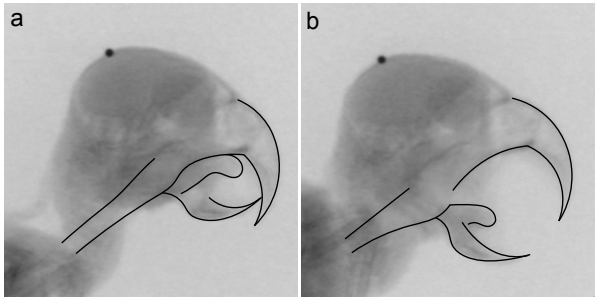

Figure 3.5. X-ray images.

This figure shows two X-ray frames of the same monk parakeet **(a)** prior to vocalizing and **(b)** during contact call production. Beak, tongue and trachea are highlighted by black lines.

Relationship between articulators and intensity

The fast FM patterns characteristic for contact calls are likely to be caused by the sound source and only marginally influenced by articulatory movements of the upper vocal tract since 1) in both contact and greeting calls tongue and beak movements as well as tracheal contraction are comparable and 2) changes in articulatory configurations are slow compared to FM. Changes in resonance patterns of greeting calls, however, are likely to be influenced by articulator movements. Unfortunately it was not possible to establish clear relationships between articulator configuration and formant changes because it is not clear how the sound source behaves in this species which therefore precludes extracting the filter characteristics. However, we detected positive correlations between articulator movements (beak opening, tongue height change and tracheal contraction) and intensity for greeting calls and chatter sounds in several birds (Fig. 3.6, Table 3.4). We did not find a correlation between beak movements and intensity and tongue height changes and intensity for contact calls, although this might be due to the fact that contact calls are generally rather loud calls and there is little variation in intensity.

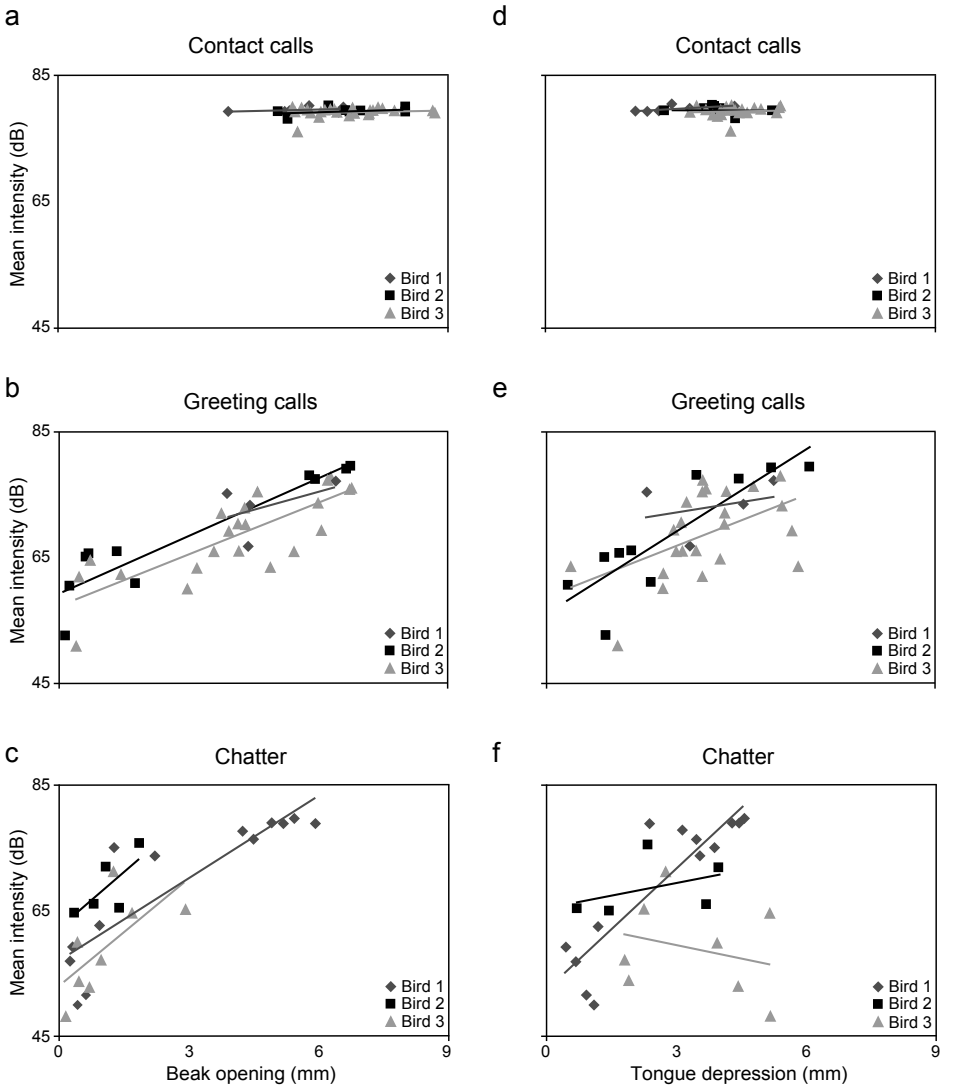

Figure 3.6. Correlations between articulator displacements and vocalization intensity.

This figure shows six scatter plots in which **(a)** beak opening and intensity for contact calls, **(d)** tongue depression and intensity for contact calls, **(b)** beak opening and intensity for greeting calls, **(e)** tongue depression and intensity for greeting calls, **(c)** beak opening and intensity for chatter sounds and **(f)** tongue depression and intensity for chatter sounds are plotted against each other for all three birds. Table 3.4 lists which of these correlations are significant. dB, decibel; mm, millimeters.

Table 3.4. Correlations between distances and intensity.

Distance measured	Vocalization		Bird 1	Bird 2	Bird 3
		Spearman's rho	0.257	0.231	-0.035
	contact call	p	0.623	0.521	0.858
		n	6	10	28
		Spearman's rho	0.400	0.915	0.816
Beak opening	greeting call	p	0.600	**< 0.01**	**< 0.01**
		n	4	10	23
		Spearman's rho	0.918	0.700	0.762
	chatter	p	**< 0.01**	0.188	**0.028**
		n	13	5	8
		Spearman's rho	0.657	-0.103	0.151
	contact call	p	0.156	0.777	0.442
		n	6	10	28
		Spearman's rho	0.400	0.867	0.532
Tongue depression	greeting call	p	0.600	**< 0.01**	**< 0.01**
		n	4	10	23
		Spearman's rho	0.813	0.600	0.286
	chatter	p	**< 0.01**	0.285	0.493
		n	13	5	8
		Spearman's rho			-0.179
	contact call	p			0.702
		n			7
		Spearman's rho			0.397
Tracheal shortening	greeting call	p			0.061
		n			23
		Spearman's rho			0.587
	chatter	p			**0.045**
		n			12

This table shows the correlations between three distances measured (beak opening, tongue depression and tracheal contraction) and mean intensity for all vocalizations measured. Intensity was measured over the whole vocalization. Significant p-values are printed bold. *n*, number of vocalizations measured.

Discussion

Our study is the first to investigate vocal tract articulation in a naturally vocalizing parrot species using X-ray cinematographic imaging. Our results demonstrate that monk parakeet vocalizations are accompanied by prominent changes in beak gape, tongue position and tracheal length. These findings are partly consistent with what has been previously reported for an African grey parrot imitating speech (Warren *et al.* 1996). While previous studies have indicated that retraction and extension of the tongue between back and front positions, respectively, seem to be particularly important to mimic human speech (Warren *et al.* 1996) and modulate formant patterns (Beckers *et al.* 2004), our results show that monk parakeets especially manipulate the high-low dimension when vocalizing while they might be able to move their tongue in a horizontal plane more than they actually do when communicating naturally. Given that monk parakeets can mimic human speech, which seems to require extensive control over the front-back position of the tongue, one wonders why they do not use this dimension as much in their own vocalizations. Nevertheless it is obvious from the videos that tongue position also changes with respect to frontedness, although it is difficult to reliably quantify these patterns.

Beak gape which has been found to correlate with frequency changes in many bird species (Hausberger *et al.* 1991; Westneat *et al.* 1993; Hoese *et al.* 2000; Podos *et al.* 2004; Goller *et al.* 2004) also changes up to 6.68 mm, in the index of beak gape used in the current study, in vocalizing monk parakeets although we could not establish a quantitative relationship with frequency patterns. However, it seems that beak gape and tongue position can change independently from each other at least to a certain degree since we observed prominent tongue movements in soft greeting calls while beak gape changed only slightly. Therefore we can conclude that tongue position is not merely incidental to beak opening, a question that arose in a previous study (Warren *et al.* 1996).

Furthermore the strong tracheal shortening which we observed on the videos provides convincing evidence for a new type of vocal articulator in birds. The shortening is accompanied by a caudal movement of the lower mandible and the hyoid skeleton and although it might be a passive process resulting from the movements of other articulators it is very likely to have an effect on the sound produced. A former study (Daley & Goller 2004) investigating tracheal length changes in singing zebra finches found that at the beginning of a song bout and between motifs tracheal length decreased. While the initial contraction was actively mediated by syringeal muscles the shortening

within the motif seemed to be the result of pressure changes in the interclavicular air sac and could not be related to frequency patterns of the song. However, length changes were small (<0.2 mm) within a song and represented only about 3 % of the length of the trachea and therefore are unlikely to have a strong effect on resonance patterns. Even within the family *Psittacidae* the degree to which the trachea can contract seems to vary noticeably between species since in African grey parrots the trachea can only stretch about 10 % (Pepperberg *et al.* 1998) while in our monk parakeets the trachea showed very little resistance to tracheal shortening. Future research will have to reveal how exactly acoustic features of vocalizations are influenced by tracheal length changes.

We also found a significant positive correlation between beak opening and amplitude in greeting calls in two of the three birds. The same significant correlation was found for tongue height change and amplitude in the greeting calls of the same birds. The reason why we did not find a correlation in one of the birds between beak movement and amplitude as well as tongue height change and amplitude for greeting calls is most probably due to the small sample size of only 4 greeting calls that were of sufficient quality for analysis (Table 3.4). The analysis revealed more positive correlations for chatter sounds in some individuals but not for contact calls, likely because contact calls were generally rather loud and showed little variation in amplitude (Fig. 3.6 a, d). These findings largely agree with earlier reports on zebra finches producing loud notes with large beak gapes (Ohms *et al.* 2010b).

Judging from the X-ray videos it seems that monk parakeets do not expand the cervical end of the esophagus to form a large OEC as do songbirds. In accordance with this observation is the fact that when obtaining silicone casts of the oral cavity from dead monk parakeets no silicone entered the esophagus while the cranial part of the trachea and the glottis were filled with silicone. Further research is needed to clarify if the esophagus contributes to vocal production in parrots at all.

Overall we have shown that monk parakeets use several articulators when producing species-specific sounds with tongue height changes, beak gape opening and tracheal length changes being the most obvious movements. However, tongue movements in the horizontal direction, although less prominent, are also likely to affect sound production while other possible articulators such as glottal opening still have to be identified. Experimentally manipulating such structures and obtaining cineradiographic data on mimicking parrots would provide further insight into the mechanisms underlying vocal production and would be of great interest for comparing the role of the tongue in human speech production and in parrot speech imitation.

Acknowledgements

We thank Amy Coy for assistance conducting the experiment, Kenneth Kragh Jensen for general discussion and Inge van Noortwijk for artwork in figure 3.1. Funding was provided by the Netherlands Organization for Scientific Research (NWO), grant Number 815.02.011 to CtC and grant NINDS R01 NS029467 from NIH to RAS. All animal procedures were reviewed and approved by the Institutional Animal Care and Use Committee and the Radiation Safety Office of Indiana University, and comply with the 'Principles of animal care', publication no. 86-23, revised 1985 of the National Institute of Health.

4

Zebra finches exhibit speaker-independent phonetic perception of human speech

Verena R. Ohms, Arike Gill, Caroline A. A. van Heijningen, Gabriël J. L. Beckers & Carel ten Cate

Humans readily distinguish spoken words that closely resemble each other in acoustic structure, irrespective of audible differences between individual voices or sex of the speakers. There is an ongoing debate about whether the ability to form phonetic categories that underlie such distinctions indicates the presence of uniquely evolved, speech-linked perceptual abilities or is based on more general ones shared with other species. We demonstrate that zebra finches (*Taeniopygia guttata*) can discriminate and categorize monosyllabic words that differ in their vowel and transfer this categorization to the same words spoken by novel speakers independent of the sex of the voices. Our analysis indicates that the birds, like humans, use intrinsic and extrinsic speaker normalization to make the categorization. This finding shows that there is no need to invoke special mechanisms, evolved together with language, to explain this feature of speech perception.

Published in Proceedings of the Royal Society Series B-Biological Sciences (2010) 277: 1003-1009.

Introduction

Human speech is a hierarchically organized coding system. A finite number of meaningless sounds, called phonemes, which are classes of speech sounds that are identified as the same sound by native speakers, are combined into an infinite set of larger units: morphemes or words. These larger units carry meaning and therefore allow linguistic communication (Yule 2006). An important role in the coding process is played by formants - vocal tract resonances that can be altered rapidly by changing the geometrical properties of the vocal tract using articulators such as tongue, lips and soft palate (Titze 2000). Changing the formant pattern of an articulation results in a different vowel produced (Fig. 4.1).

It has been argued in the past that many characteristics of speech are uniquely human (e.g. Lieberman 1975, 1984). Therefore it was a revolutionary finding when Kuhl and Miller (1975, 1978) who tested chinchillas on their ability to discriminate between /d/ and /t/ consonant-vowel syllables found that these animals have the same phonetic boundaries as humans and thereby challenged the view that the mechanisms underlying speech perception are uniquely human. A few years later the same phonetic boundary effect has been shown in macaques (Kuhl & Padden 1982). Nevertheless, there is still an ongoing debate about which parameters of human speech production and perception are unique to humans, with the implication that they evolved together with speech or language, and which are shared with other species (Liberman & Mattingly 1985; Hauser *et al.* 2002; Trout 2003; Diehl *et al.* 2004; Pinker & Jackendoff 2005).

One of the most important phenomena in human speech concerns our ability to recognize words regardless of individual variation across speakers. Although human voices differ in acoustic parameters such as fundamental frequency and spectral distribution, we are able to distinguish closely similar words by using the relative formant frequencies in dependence of the fundamental frequency of an utterance. This feature enables the intelligibility of speech (Nearey 1989; Fitch 2000; Assmann & Nearey 2008). But does this mean that the human ability to perceive and normalize formant frequencies in order to develop an abstract formant percept evolved together with speech and language? Or has the evolution of language exploited a pre-existing perceptual property that allowed formant normalization? An important way to test this question is by examining whether this feature is present in other animals. If so, this suggests that it is not a uniquely evolved faculty.

Here we examined whether zebra finches trained to distinguish two words differing in one vowel only and produced by several same-sex speakers, generalize the distinction to a novel set of speakers of (1) the same sex and (2) the opposite sex. We chose for natural human voices instead of artificial stimuli to confront the animals with a situation humans have to deal with every day when vocally communicating: extracting the relevant sound features from irrelevant ones while listening and building up a percept that allows categorization of these words when originating from novel voices.

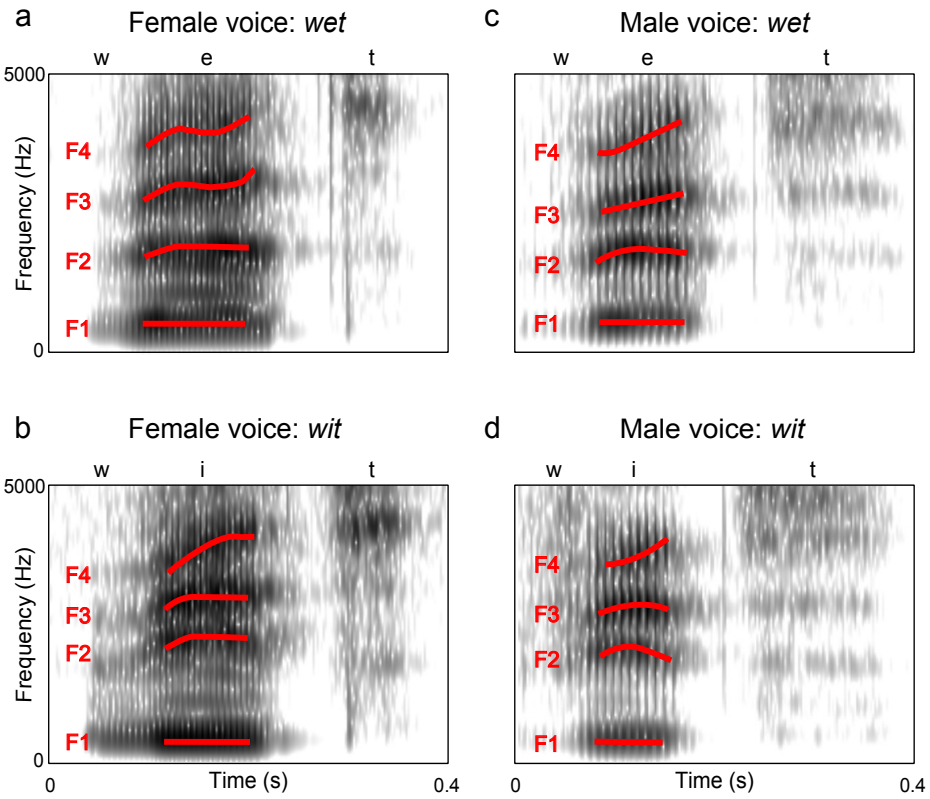

Figure 4.1. Spectrograms of human voices

(a) Female voice saying *wet*; **(b)** Female voice saying *wit*; **(c)** Male voice saying *wet*; **(d)** Male voice saying *wit*. Red lines indicate the formant frequencies. Note the difference in the distance of the first and second formant frequencies. In a this distance is smaller than in b and the same applies for c and d. F1, first formant; F2, second formant; F3, third formant; F4, fourth formant; s, seconds; Hz, Hertz.

Material and Methods

Subjects

We used three male and five female zebra finches (*Taeniopygia guttata*, aged 6 months to 2 years) from the Leiden University breeding colony. Prior to the experiment, birds were housed in groups of two or three animals and were kept on a 13.5 L:10.5 D schedule. Food, grit, and water were provided *ad libitum*. None of the birds had previous experience with psychophysical experiments. At the beginning of the study every animal was weighed to allow monitoring of the nutritional state. During the experiment the amount of food eaten by the birds was checked daily. If an animal ate less than necessary it was provided with additional food. In this case the bird was also weighed to ensure that it did not lose more than 20% of its initial body weight. All animal procedures were approved by the animal experimentation committee of Leiden University (DEC number 08054).

Stimuli

We obtained naturally spoken Dutch words from second year students at Leiden University. A total of 10 female and 11 male native speakers of Dutch were recorded in the phonetics laboratory of the Faculty of Humanities, Leiden University using a Sennheiser RF Condenser Microphone MKH416T and Adobe Audition 1.5 software with 44.1 kilosamples/s, at a 16 bit resolution. Every speaker was asked to read a list of Dutch words in which the stimuli '*wit*' (wIt) and '*wet'* (wɛt) were embedded to prevent list-final intonation effects. The recordings were processed afterwards using the software Praat (version 4.6.09) freely available at www.praat.org (Boersma, 2001) by cutting out the words *wit* and *wet* and saving both as separate wave files for each voice. To prevent intensity differences between stimuli from playing a role in the discrimination process, the average amplitude of all female and male voices respectively was normalized by using the root mean square of the average acoustic energy and equalizing it. During the experiment all stimuli were played back at approximately 70 dB SPL(A).

Apparatus

The experiment was conducted in a Skinner box described earlier (Verzijden *et al.* 2007), which was placed in a sound attenuated chamber. Sounds were played through a Vifa MG10SD-09-08 broadband loudspeaker at approximately 70 dB SPL(A) attached one meter above the Skinner box. A fluorescent lamp (Lumilux De Luxe Daylight, 1150 lm,

L 18 W/ 965, Oscam, Capelle aan der IJssel, The Netherlands) served as light source and was placed on top of the Skinner box. It was switched on automatically every day from 7:00 a.m. to 8.30 p.m. whereby the light was gradually increasing and decreasing in a 15 minutes time window at the beginning and end of the light cycle respectively.

Discrimination learning

To train the birds to discriminate between acoustic stimuli we used a 'Go/NoGo' operant conditioning procedure (Verzijden *et al.* 2007). The positive ('Go') stimulus (S^+) was an average zebra finch song whereas the negative ('NoGo') stimulus (S^-) was a pure tone of 2 KHz constructed in Praat (Boersma 2001). During the training the birds had to learn that responding to S^+ would lead to a 10 second food reward with access to a commercial seed mix, whereas responding to S^- would cause a 15 second punishment interval with the lights in the experimental chamber going out (Fig. A 4.1).

Experiment

The actual experiment consisted of four successive phases. As soon as the birds reached the discrimination criterion (*d'*=1.34) which we defined as a high response rate to the Go-stimulus (75% or more) and a low response rate to the NoGo-stimulus (25% or less) over three consecutive days they were transferred to the next stage. During the first stage of the experiment every bird had to learn to discriminate the words *wit* and *wet* of a single person (stage 1) whereby every bird started with a different voice. Four groups with two birds per group were formed (Fig. A 4.2). Two groups started with female voices and the other two groups with male voices. One of the groups that began the experiment with a female voice received *wit* as positive and *wet* as negative stimulus and vice versa for the other group. The birds that started with the male voices were treated accordingly. After the birds had reached the discrimination criterion they were switched to the next stage (stage 2) in which four new minimal pairs of the same sex as the first voice were added. After reaching the discrimination criterion birds were transferred to stage 3 in which the five voices used in stage 2 were replaced by five new voices of speakers of the same sex. In the final stage of the experiment (stage 4) the birds were confronted with five new voices of the opposite sex. The experiment was finished after the birds again fulfilled the discrimination criterion. To prevent pseudoreplication voices were randomly balanced over the four groups.

Performance evaluation

To assess discrimination performance between *wit* and *wet* we calculated the *d'* and 95% confidence interval following the procedure used and described by others (Macmillan & Creelman 2005; Gentner *et al.* 2006) for every bird for the first 100, 200 and 300 trials directly after transition between the different phases. This is a sensitivity measure that subtracts the *z* score of the false-alarm rate (F), which is defined as the proportion of responses to a NoGo-stimulus divided by the total number of NoGo-stimulus presentations, from the *z* score of the hit rate (H), which is the proportion of responses to a Go-stimulus divided by the total number of Go-stimulus presentations. This measure allows evaluating how well two stimuli are discriminated from each other: $d' = z$ (H) – z (F). A *d'* of zero indicates no discrimination whereas a lower bound of the 95% confidence interval above zero can be considered to indicate significant discrimination (Macmillan & Creelman 2005; Gentner *et al.* 2006). Moreover this measurement is unaffected by a potential response bias (Macmillan & Creelman 2005).

Acoustic measurements

In order to detect acoustic features that might have enabled distinction between *wit* and *wet* we measured word and vowel duration as well as fundamental frequency and the mean first (F1) and second (F2) formant frequencies of both words obtained by the different speakers using Praat software (Boersma 2001). We ran two-tailed Wilcoxon signed ranks tests separately for male and female voices to detect significant differences of the acoustic characteristics between *wit* and *wet*.

Results

In the first phase of the experiment all birds learned to discriminate reliably between the two words *wit* and *wet* and fulfilled the discrimination criterion after an average of 41 blocks (40.72 ± 3.41 s.e.m.) with 100 trials per block.

However, this outcome does not imply generalized categorical discrimination as the birds might have learned the individual features of the training stimuli. In order to show that the birds had developed a generalized percept their performance should be independent of individual voices. In the next phase we therefore added four additional minimal pairs recorded by same-sex speakers to the first stimulus pair but maintained

the same learning criterion. The mean *d'* (which is a measure of how well two stimuli are discriminated from each other) of the first 100 trial block after this transition was 0.77 ± 0.30 (*d'* ± s.e.m.) which is clearly above chance level (*d'*= 0). After transition of stimulus sets (Fig. 2b) five out of eight birds immediately performed above chance level and all birds achieved a significant performance within the first three blocks after transition (mean *d'*= 0.94 ± 0.17 s.e.m. with the lower bound of the 95% confidence interval ranging from 0.14 to 0.94).

It could be argued that these results are biased through the incorporation of an already familiar voice in the stimuli sets. Hence, in the subsequent phase we switched to five completely unknown speakers of the same sex (Fig. 4.2c). Again, the average *d'* was already highly above chance level over the first 100 trials after transition (*d'*= 1.01 ± 0.32 s.e.m.) for six out of eight birds. Within 300 trials after transition all birds showed clear discrimination with a lower bound of the 95% confidence interval ranging from 0.2 to 1.57. Thus, the birds seem to have formed a generalized percept.

So far all voices were of the same sex and overlapped in several features. Therefore a more critical test is to check whether the birds are able to transfer the discrimination to the same words spoken by the opposite sex, i.e. whether the relevant acoustic features can be transferred to a context with larger differences in pitch and timbre compared to voices within the same sex. Consequently, we switched to five new voices of the opposite sex in the last phase of the experiment (Fig. 4.2d). This time all birds discriminated well above chance level (average *d'*= 0.9 ± 0.59 s.e.m.) within the first block after transition with the lower bound of the 95% confidence interval ranging from 0.02 to 0.59.

We measured various acoustic characteristics that may have allowed discrimination (Table A 4.1). It is possible that a consistent difference in either vowel or word duration between *wit* and *wet* enabled distinction, but neither vowel nor word duration differed regarding the male voices. There was a significant difference in vowel duration for the female voices (Wilcoxon signed ranks test, n = 10, T+ = 47, T- = 8, p = 0.048) with /I/ being shorter than /ɛ/, but as all birds showed a generally high selectivity irrespective of the sex of the voices it can be assumed that vowel duration was not involved in discrimination. Another cue that might have influenced discrimination is the fundamental frequency of the voices which is known to differ between vowels with /ɛ/ having a slightly lower fundamental frequency than /I/ (Peterson & Barney 1952). This observation complies with our measurements although the difference is only significant for the male voices (Wilcoxon signed ranks test, n = 11, T+ = 59.5, T- = 6.5, p = 0.018). However, the disparity in fundamental frequency between voices is much larger than

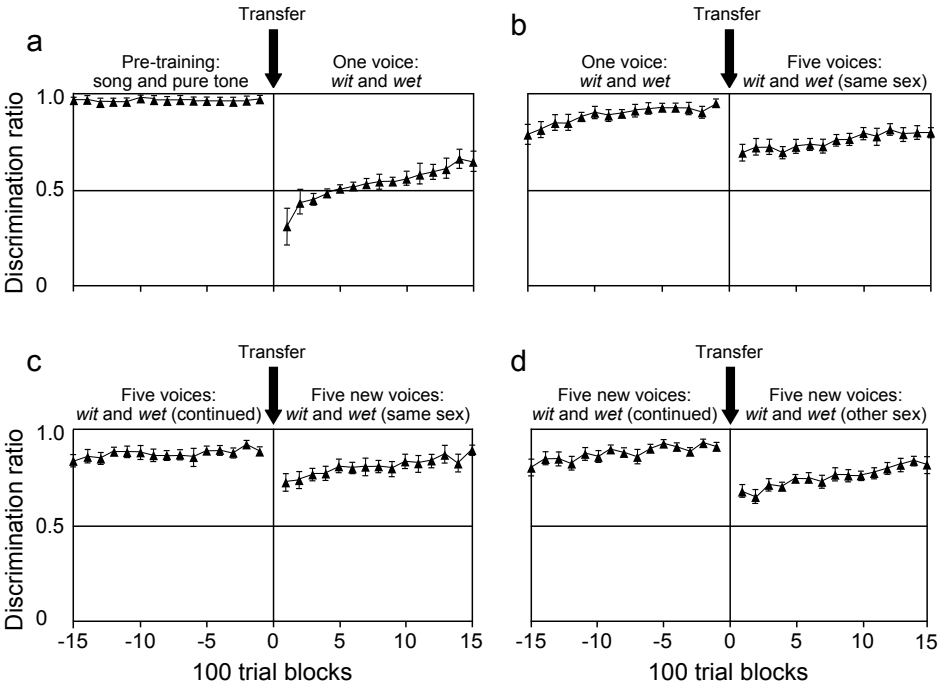

Figure 4.2. Transitions between discrimination stages.

Displayed is the discrimination ratio of the last fifteen 100 trial blocks before and after a transition between two stages. A discrimination ratio of 1.0 reflects perfect discrimination whereas a discrimination ratio of 0.5 indicates chance performance. The discrimination ratio is calculated as follows: (Go S+/ total S+) / [(Go S+/ total S+) + (Go S-/ total S-)]. **(a)** Transition between the pre-training phase in which all birds had to discriminate a zebra finch song from a 2 kHz pure tone and the training phase in which the animals were confronted with the first minimal pair. **(b)** shows the transition between the training phase and the subsequent experimental stage in which four additional minimal pairs of the same sex were added to the already familiar voice. **(c)** Transition between minimal pairs of now five familiar voices and five completely unknown voices of the same sex. **(d)** Transition from five voices to five new voices of the other sex. kHz, Kilohertz; Go S+, number of responses to a positive stimulus; total S+, number of positive stimulus presentations; Go S-, number of responses to a negative stimulus; total S-, number of negative stimulus presentations.

within voices, so that this feature alone cannot be sufficient for discrimination.

On the other hand we found a highly significant difference in the formant frequencies of the first (F1) and second (F2) formant between *wit* and *wet* as expected (Fig. 4.3a, Table A 4.1 and Table A 4.2). However, if the birds had only paid attention to the absolute frequency of F1 they should have treated the female *wit* as the male *wet* due to the overlap in F1 frequency (Fig. 4.3a, Table A 4.2) whereas in case they based their discrimination on F2 only they should have treated the male *wit* as the female *wet* as these words overlap in F2 frequency (Fig. 4.3a, Table A 4.2).

From phonetic research we know that humans do not discriminate vowels solely based on their absolute formant frequencies, but rather rely on relative formant ratios in dependence of the fundamental frequency (F0) of a speaker (Assmann & Nearey 2008). A common way of illustrating the relationship between formant frequencies and fundamental frequency as a method of intrinsic speaker normalization (Magnuson & Nusbaum 2007) is plotting the difference between F0 and F1 against the difference of F1 and F2 in Bark (Fig. 4.3b), which can be regarded as a two-dimensional perceptual similarity measure of different sounds. Applying this method to our stimuli, results in two clearly separate vowel categories despite an extensive overlap between the sexes (Fig. 4.3b).

Discussion

Previous studies on speech perception by non-human animals have suggested that the ability to discriminate human speech sounds based on their formant patterns, such as demonstrated in our study, is not unique to humans, but can be found in other taxa as well. Such studies have been carried out in several mammals, e.g. cats, chinchillas, monkeys and rats (Burdick & Miller 1975; Kuhl & Miller 1975, 1978; Kuhl 1981; Hienz & Brady 1988; Hienz *et al.* 1996; Eriksson & Villa 2006), and birds, such as budgerigars, pigeons, red-winged blackbirds and quail (Hienz *et al.* 1981; Kluender *et al.* 1987; Dooling *et al.* 1989; Dooling & Brown 1990; Dent *et al.* 1997). Most of these experiments used synthesized speech sounds lacking natural variation (Kuhl & Miller 1978; Hienz *et al.* 1981; Kuhl 1981; Hienz & Brady 1988; Dooling *et al.* 1989; Hienz *et al.* 1996; Dent *et al.* 1997; Eriksson & Villa 2006) to demonstrate that the way in which these were discriminated and categorized is equivalent to how humans do so. However, in order to show that animals do use the same mechanisms as humans do when categorizing speech

sounds it is crucial to work with natural and varying stimuli, which has been done only in a minority of studies (Burdick & Miller 1975; Kuhl & Miller 1975; Kluender *et al.* 1987; Dooling & Brown 1990). However, these studies either used isolated vowels or speech sounds from a small number of speakers. While definitely instructive none of these studies fulfilled the requirements of testing a phonemic contrast by employing different vowels embedded in a minimal pair of words. This might seem to be a minor detail when studying speech perception by animals, but yet is essential as humans do not simply make one-bit discriminations between single phonemes (Pinker & Jackendoff 2005), but have to extract relevant information from words that closely match each other in acoustic structure in other respects. Furthermore it is indispensable to use sufficiently different speakers (Magnuson & Nusbaum 2007).

Our experiment controlled for the above mentioned factors and our results strongly suggest that zebra finches use formants to make phonetically relevant discriminations and, similar to humans, abstract away from irrelevant variation between voices.

For humans, 'intrinsic normalization' theories (Nearey 1989) account for the phenomenon that sounds which are perceived as one phoneme can have several acoustic realizations (Liberman *et al.* 1967) by constituting that every speech sample can be categorized using a normalizing transformation. Our analyses indicate that zebra finches use a similar mechanism. However, these theories cannot explain the learning process also revealed by our data. Although the birds were able to immediately categorize *wit* and *wet* independent of speaker variability their performance dropped when confronted with new voices and then improved constantly (Fig. 4.2). Experiments with humans have also shown a clear speaker effect on speech discrimination. In a study by Magnuson and Nusbaum (2007) human subjects were presented with orthographic forms of a target vowel on a computer screen and asked to press the space bar when they heard the target vowel that they saw on the screen. Every subject had to do this task under different conditions, namely 'blocked-talker' condition, which means that all stimuli were from the same talker and 'mixed-talker' condition, which means that the stimuli were from two different talkers. In most cases the response time was significantly higher in the 'mixed-talker' condition compared to the 'blocked-talker' condition while the hit rate was significantly lower. The same speaker effect has been demonstrated by other studies in which the human ability to recognize whole words under varying conditions has been tested (Creelman 1957; Mullennix *et al.* 1989). In addition, also human subjects improve their discrimination performance over trial blocks (Mullennix *et al.* 1989) just as the

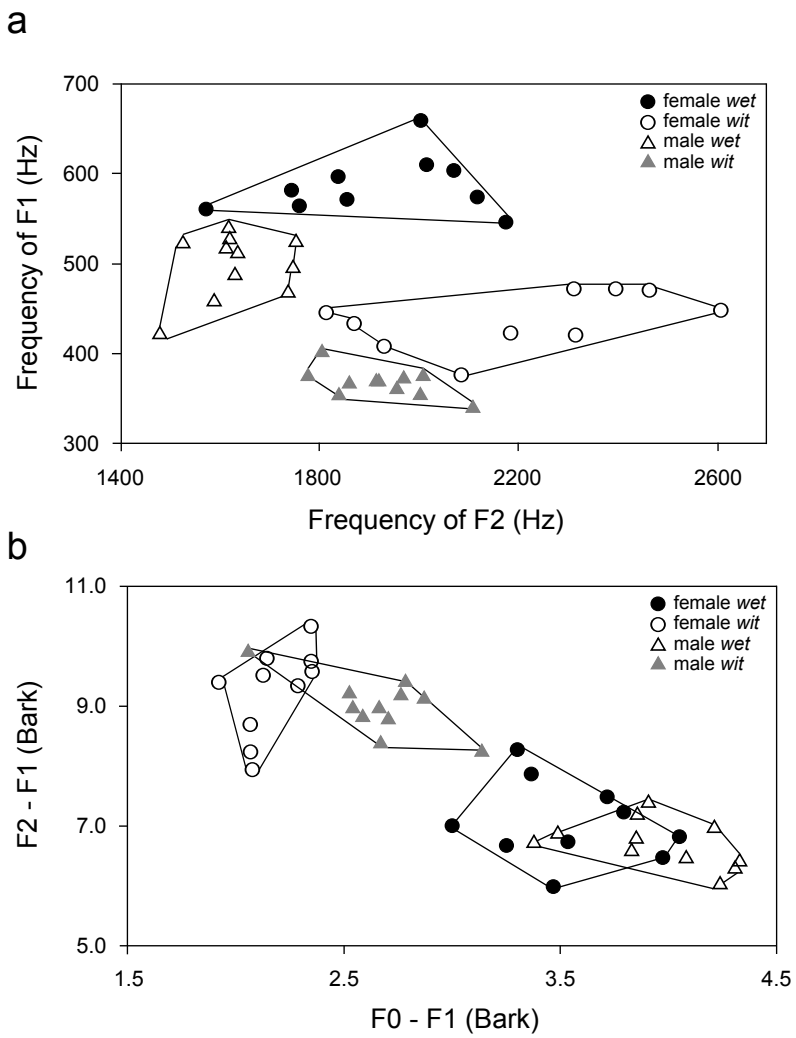

Figure 4.3. Vowel diagramms.

(a) The frequencies of the first and second formants of all individual voices saying *wit* and *wet* are plotted against each other. Especially with regard towards the second formant frequencies the male voices form denser clusters than the female voices, which show more variation. Nevertheless, the vowels /I/ and /ɛ/ can be clearly separated from each other. **(b)** The difference between the fundamental frequency and the first formant (in Bark) is plotted against the difference between the first and the second formant (in Bark) for all recordings used in the experiment. In contrast to the formant scatter plot in (a) this figure represents a two-dimensional perceptual concept in which male and female voices clearly overlap, whereas the two vowels /I/ and /ɛ/ are fully separated. F0, fundamental frequency; F1, first formant; F2, second formant.

zebra finches in the current study. This outcome indicates the presence of extrinsic normalization in humans and zebra finches, i.e. establishing a reference frame from the vowel distribution of the various speakers as a function of learned formant ranges (Magnuson & Nusbaum 2007; Nearey 1989).

So, due to the design and the results of our study our evidence holds out against arguments that in the past allowed doubts about the universality of the auditory mechanisms underlying speech perception. With respect to speaker normalization our experiment therefore provides very strong evidence that non-human animals use the same perceptual principles as humans do when discriminating speech sounds by employing a combination of intrinsic and extrinsic speaker normalization and thereby suggests that the underlying mechanisms originally emerged in a context independent of speech.

It is mainly due to the lowering of the larynx that humans can produce so many distinct speech sounds (Lieberman *et al.* 1969). However, another effect of a lowered larynx is to increase the length of the vocal tract which causes a decrease of formant frequencies. This in turn can be used to exaggerate size, and playback experiments in red deer which possess a lowered larynx too, have shown that stags respond more to roars with lower formant frequencies compared to roars with higher formants (Reby *et al.* 2005). In humans, formant frequencies are used to correctly estimate age (Collins 2000) and they strongly influence the perceived height of a speaker (Smith & Patterson 2005) and hence can serve as indexical cues next to their function of coding linguistic information. Rhesus monkeys use formants in species-specific vocalizations as indexical cues as well (Ghazanfar *et al.* 2007) and although not many studies have investigated similar phenomena in bird vocalizations it has been shown that whooping cranes for example can perceive changes in formant frequencies in their own species calls and exhibit a different response pattern to calls with higher formants compared to lower formants (Fitch & Kelley 2000). These results led to the speculation that formant perception originally emerged in a wide range of species to assess information about the physical characteristics of conspecifics and that human speech has exploited the already existing sensitivity for formant perception (Ghazanfar *et al.* 2007; Fitch 2000).

It can, of course, not be ruled out completely that unique perceptual abilities to facilitate speech perception did evolve in humans, or that the observed abilities evolved separately in birds and humans. In the latter case, this would indicate a remarkable convergence. However, our results, in combination with earlier findings, also support the hypothesis that the evolution of the variety of speech sounds in humans might have been

shaped by pre-existing perceptual abilities rather than being the result of co-evolution between the mechanisms underlying the production and perception of speech sounds.

Acknowledgements

We thank Vincent J. van Heuven for advice considering the recording of the stimuli, for permission to use the phonetics laboratory and him, Katharina Riebel, Hans Slabbekoorn and Willem Zuidema as well as two anonymous referees for useful comments on the manuscript. Funding was provided by the Netherlands Organisation for Scientific Research (NWO). Grant number 815.02.011.

Appendix

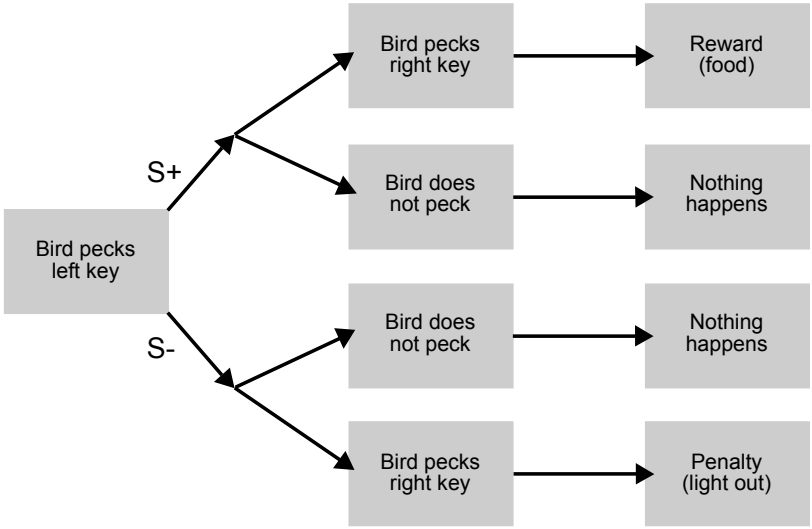

Figure A 4.1. Schematic overview of the Go/No-Go procedure.

In the Go/NoGo operant conditioning task the bird elicits every trial by pecking on the left pecking key and subsequently hears either a positive or a negative stimulus. If the bird hears a positive stimulus and pecks on the right key it receives a 10 second food reward. On the other hand if the bird responds to a negative stimulus the lights in the experimental chamber will go out for 15 seconds which serves as a punishment. If the bird does respond neither to a positive nor a negative stimulus within 6 seconds after it heard the sound nothing happens and a new trial can be elicited by the bird by pecking the left key again. S+. positive stimulus; S-, negative stimulus.

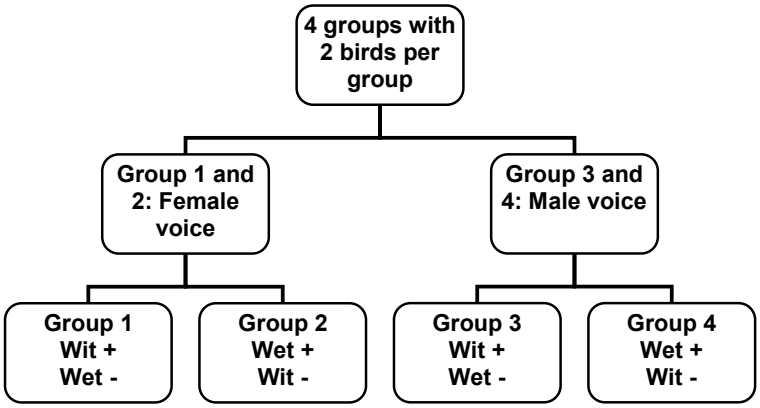

Figure A 4.2. Overview over the 4 different treatment groups.

Groups 1 and 2 started the discrimination experiment with a female voice, whereas groups 3 and 4 started with a male voice. For groups 1 and 3 *wit* was the positive stimulus and *wet* the negative stimulus, for groups 2 and 4 *wet* was the positive and *wit* the negative stimulus.

Table A 4.1. Results of the statistical voice analysis.

			'wit'	*'wet'*			
Male voices	**Parameter**	***n***	***x* ± *s.d.***	***x* ± *s.d.***	***T+***	***T-***	***p***
	Word duration (ms)	*11*	415 ± 33.5	399 ± 43.4	47	19	0.240
	Vowel duration (ms)	*11*	117 ± 15.6	121 ± 14.4	21.5	33.5	0.541
	F0 (Hz)	*11*	123 ± 14.8	115 ± 10.4	59.5	6.5	**0.018**
	F1 (Hz)	*11*	369 ± 15.7	500 ± 35.8	66	0	**0.001**
	F2 (Hz)	*11*	1923 ± 99.1	1631 ± 87.9	66	0	**0.001**
	F1/F2 ratio	*11*	0.193 ± 0.016	0.307 ± 0.023	0	66	**0.001**
Female voices	**Parameter**	***n***	***x* ± *s.d.***	***x* ± *s.d.***	***T+***	***T-***	***p***
	Word duration (ms)	*10*	427 ± 59.8	439 ± 87.0	24	31	0.721
	Vowel duration (ms)	*10*	109 ± 19.6	114 ± 22.0	47	8	**0.048**
	F0 (Hz)	*10*	220 ± 19.8	213 ± 20.1	46	9	0.064
	F1 (Hz)	*10*	437 ± 31.2	587 ± 32.1	0	55	**0.002**
	F2 (Hz)	*10*	2199 ± 267.8	1916 ± 192.1	54	1	**0.004**
	F1/F2 ratio	*10*	0.201 ± 0.023	0.309 ± 0.031	0	55	**0.002**

Wilcoxon signed ranks test (two-tailed) were conducted to calculate differences in various acoustic parameters between the words *wit* and *wet* whereby male and female voices were compared separately. Significant p-values are printed bold. ms, milliseconds; Hz Hertz; F0, fundamental frequency; F1 first formant; F2 second formant.

Table A 4.2. Formant values of all used voices.

Female						Male					
'wit'			*'wet'*			*'wit'*			*'wet'*		
F1 (Hz)	**F2 (Hz)**	**ratio**	**F1 (Hz)**	**F2 (Hz)**	**ratio**	**F1 (Hz)**	**F2 (Hz)**	**ratio**	**F1 (Hz)**	**F2 (Hz)**	**ratio**
470	2466	0.191	659	2006	0.329	341	2108	0.162	497	1748	0.284
422	2186	0.193	596	1839	0.324	363	1955	0.186	528	1620	0.326
448	2608	0.172	546	2176	0.251	369	1860	0.198	519	1612	0.322
376	2086	0.180	582	1744	0.334	356	1839	0.194	489	1629	0.300
472	2397	0.197	609	2017	0.302	374	1969	0.190	526	1753	0.300
471	2314	0.204	574	2119	0.271	356	2003	0.178	470	1737	0.271
408	1929	0.212	561	1573	0.357	403	1804	0.223	525	1524	0.344
420	2316	0.181	603	2070	0.291	377	2008	0.188	424	1478	0.287
445	1816	0.245	571	1858	0.307	370	1912	0.194	542	1616	0.335
433	1870	0.232	565	1759	0.321	371	1916	0.194	514	1634	0.315
						377	1775	0.212	460	1588	0.290

In this table the frequencies of the first and second formant and the values of the first formant divided by the second formant are listed for all individual voices. F1, first formant; F2, second formant; Hz, Hertz.

5

Zebra finches and Dutch adults exhibit the same cue weighting bias in vowel perception

Verena R. Ohms, Paola Escudero, Karin Lammers, & Carel ten Cate

Vowels in human speech differ from each other in several acoustic features. A major question in speech perception concerns which of these features are critical to distinguish different vowels, i.e. whether features are weighted differently ('acoustic cue weighting'). Human infants for instance, are more sensitive to low frequency components when discriminating vowels, but it is unclear whether adults are too. Also, while animals are known to perceive speech sound contrasts, it is unknown if they exhibit a cue weighting bias, or if this is a uniquely human trait, linked to using speech. We provided zebra finches (*Taeniopygia guttata*) and human adults with the same task of discriminating words that had incorporated vowels which differed and overlapped in several frequency components. We show that they both exhibit a highly similar acoustic cue weighting bias. In contrast to human infants, however, both zebra finches and human adults pay more attention to high frequency components. Our results demonstrate that cue weighting in speech perception is not a uniquely human characteristic and thus need not be closely linked to experience with speech in general or with vowels in particular. We suggest that both humans and zebra finches are born with specific perceptual biases, which at least for humans might shift developmentally, perhaps as a result of their acoustic environment.

Manuscript

Introduction

The evolution of speech and language is still a fiercely debated topic among scientists from various disciplines (Hauser *et al.* 2002; Pinker & Jackendoff 2005; Anderson 2008; Fitch 2010). The original assumption that 'speech is special' (Liberman 1982) and that the mechanisms underlying speech perception are uniquely human (Lieberman 1975) has been challenged over the years by numerous studies indicating that the general ability of speech perception is widely shared with other species including both mammals (Kuhl & Miller 1975; Hienz *et al.* 1996; Eriksson & Villa 2006) and birds (Kluender et al. 1987; Dooling & Brown 1990; Ohms *et al.* 2010a). The categorical perception of speech sounds previously thought to be uniquely human is just as present in other animals (Kuhl & Miller 1975; Kluender *et al.* 1987) as is the capacity for vocal tract normalization (Ohms *et al.* 2010a). However, most research treated speech sounds as unimodal entities and has not considered the fact that multiple acoustic features are involved in their production and perception.

A major unsolved question in speech perception for humans as well as other species regards the relative contribution that those different acoustic features have in the perception of speech sound contrasts. Vowels are characterized by at least two types of formant frequencies. Formants are vocal tract resonances that are not present in most consonants (Titze 2000). The frequency values of formants vary between vowels and are dependent on the position of the tongue in the mouth cavity. For instance, the vowels in the syllables /dIt/ and /dut/ have different first (F1), second (F2) and third formant (F3) frequency values: /u/ has lower F1, F2 and F3 values than /I/ (Fig. 5.1).

Studies with human infants (Lacerda 1993, 1994; Curtin *et al.* 2009) have shown that both Swedish and Canadian-English babies perceive low formant frequencies, i.e. F1 differences, more readily than high formant frequencies, i.e. differences in F2 and F3, when distinguishing syllables that differ only in their vowel sounds. The authors of the last study explain this as a result of Canadian-English having more vowels which differ more in F1 than in F2 values. Thus, 15-month-old infants seem to exhibit a cue weighting bias towards the acoustic feature that is most important in their native language, a finding that is compatible with the fact that infants start to discriminate only the vowels of their native language, and not those of other languages, by their sixth month of age (Polka & Werker 1994). Interestingly, however, infants aged 3 to 12 months whose native language has more vowels that differ in high formant frequencies than English, namely Swedish, are also better at discriminating F1 than F2 differences

(Lacerda 1993, 1994), which suggests a universal human bias towards lower frequencies in vowel perception. However, to date it remains unclear if these biases are strictly linked to speech sound perception and hence form a uniquely human property that might change developmentally as a result of phonetic experience or not. So far, acoustic cue weighting has not been attested in any other species.

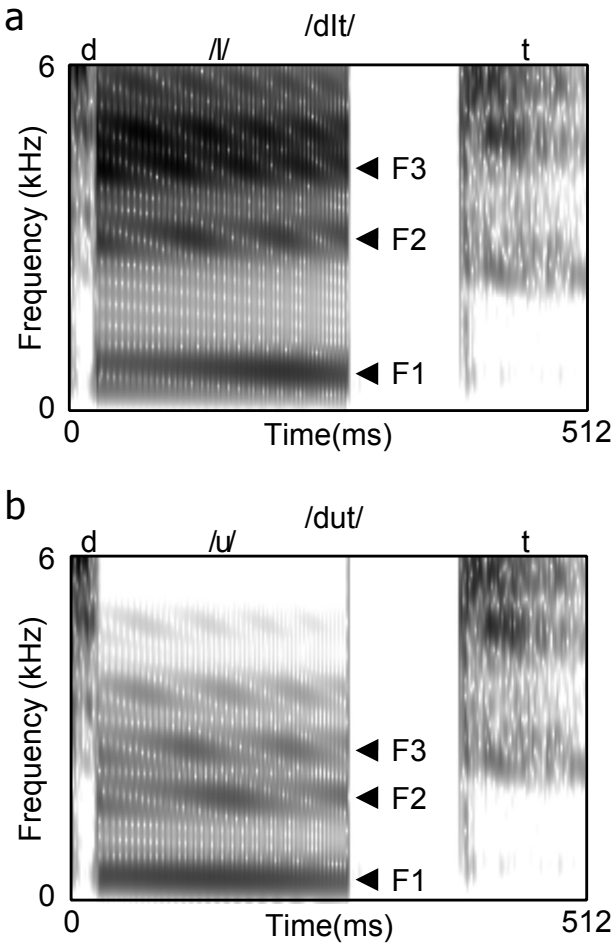

Figure 5.1. Spectrograms of two syllables differing only in their vowels.

This figure shows spectrograms of two synthetic syllables: /dIt/ and /dut/. It is clearly visible that formant frequencies differ between the vowels with lower formant frequencies in /u/ compared to /I/. F1, first formant; F2, second formant; F3, third formant; kHz, kilohertz; ms, milliseconds.

In the current study we used a Go/NoGo operant conditioning paradigm to test acoustic cue weighting in a species assumed to perceive vowel formants in similar ways as humans, namely the zebra finch (*Taeniopygia guttata*) (Ohms *et al.* 2010a; Dooling *et al.* 1995). In their own vocalizations zebra finches show a variety of note types, covering a wide frequency range, which are produced using various articulators (Ohms *et al.* 2010b). Despite similarities in vowel perception, the auditory system of a zebra finch has not been fine-tuned to the perception of human speech and it lacks experience with a particular language. Thus, one can predict that zebra finches utilize high formant frequencies more easily due to an increased sensitivity between approximately 1 and 4 kilohertz (Dooling 2004). Alternatively, existing evidence for a universal formant perception bias towards lower frequencies might transfer to zebra finches because of their human-like perception of vowels. On the other hand it still has to be explored if and how a cue weighting bias in human adults will manifest itself. Although it has been shown that Swedish babies younger than 12 months have the same cue-weighting bias as Canadian-English babies, it has yet to be shown whether extensive experience with Swedish or another language with more F2 and F3 vowel contrasts either makes both cues equally relevant, changes the bias towards higher frequencies, or does not alter the bias in human listeners. Therefore we also tested acoustic cue weighting in vowel perception in Dutch speaking adults using the same stimuli and a highly similar testing procedure as we used for the birds to make the results greatly comparable.

We used four synthetic tokens of each of the vowels /i/, /I/, /u/ and /U/ (Fig. 5.2 and Table A 5.1), which had similar F1 and F2 values to those reported earlier (Curtin *et al.* 2009). Both zebra finches and humans were trained to discriminate two of the four syllables which differed in all formant frequencies following a Go/NoGo paradigm. One syllable was associated to positive feedback, the other to negative feedback (Table 5.1). After subjects had learned to reliably discriminate between the two syllables the remaining two syllables were introduced as probe sounds. Probe sounds were never reinforced and either had the same F1 frequency as the positive stimulus and the same F2 and F3 frequencies as the negative stimulus or the other way around (Fig. 5.2 and Table A 5.1). The responses of birds and humans to the probe sounds allowed us to draw conclusions about how these sounds were perceived by the subjects.

Material and Methods

Stimuli

We used the software Praat (Boersma 2001) version 4.6.09 freely available at www.praat.org to generate four synthetic tokens of the syllables used before (Curtin *et al.* 2009) namely "deet" (/dit/), "dit" (/dIt/), and "doot" (/dut/). We also synthesized the syllable "dut" (/dUt/) to complete the set of Canadian-English high vowels. F1 and F2 values of these tokens are shown in Table A 5.1. In order to compare the use of F1 and F2 differences in vowel perception, the tokens for the contrasts /dit/-/dIt/ and /dut/-/dUt/ differed in their F1 values, while the tokens for the contrasts /dit/-/dut/ and /dIt/-/dUt/ differed in their F2 values. In terms of F1, the fourth token of each syllable, namely dit4, dIt4, and dut4, had values that fell within one standard deviation of those reported earlier (Curtin *et al.* 2009), while those for dUt4 where identical to dIt4 for F1 and to dut4 for F2. Tokens 1-4 where generated in order to examine whether variation in F1 and F2 values would lead to a different pattern in the use of these dimensions. Listeners heard only one set of tokens, e.g. /dit/1, /dIt/1, /dut/1, and /dUt/1. All synthesized vowel tokens were spliced in the middle of the same natural d_t frame, which was taken from one of the naturally produced /dut/ tokens of the study mentioned earlier (Curtin *et al.* 2009). The vowels had the same fundamental frequency (F0) and duration. They had a falling F0 contour which started at 350 Hz at the vowel onset and fell down to 250 Hz at the vowel offset, with both values being similar to those of a natural female voice. The vowels had the same duration, namely 250 ms, in order for listeners to only use vowel formant differences when discriminating the vowels in the stimuli. The vowels also differed in their F3 values because, in English, vowels with low F2 values, namely back vowels, are always produced with a low F3 value, which gives them their characteristic "rounding" feature. Thus, the script that was used to synthesize the vowels computed F3 values following the formula: F3 = F2 + 1000 Hertz for /i/ and /I/ and the formula F3 = F2 + 400 Hertz for /u/ and /U/.

Zebra finch testing

An extensive description of the testing procedure can be found elsewhere (Ohms *et al.* 2010a). Briefly, eight zebra finches were trained in a Go/NoGo operant conditioning chamber to discriminate between two syllables that differed in all formant frequencies from each other whereby every bird got a different set of stimuli (Table 5.1). One of the syllables was associated to positive feedback, the other to negative feedback (Table 5.1).

Each trial was initiated by the birds pecking a report key which resulted in playback of either the positive or the negative stimulus. The birds had to peck a response key after hearing the positive stimulus (S+), e.g. /dit/, in order to get a food reward while ignoring the negative stimulus (S-), e.g. /dUt/. Responding to the negative stimulus caused a 15 seconds time out in which the light in the experimental chamber went out. Playback of the positive and negative stimulus was randomized with no more than three consecutive positive or negative stimulus presentations. After each bird had reliably learned to discriminate between the two syllables the remaining two syllables were introduced as probe sounds in 20% of the trials. Probe sounds were never reinforced and either had the same F1 frequency as the positive stimulus and the same F2 and F3 frequencies as the negative stimulus or the other way around (Fig. 5.2 and Table A 5.1). The responses of the birds to the probe sounds allowed us to draw conclusions about how these sounds were perceived by the birds. All animal procedures were approved by the animal experimentation committee of Leiden University (DEC number 09058).

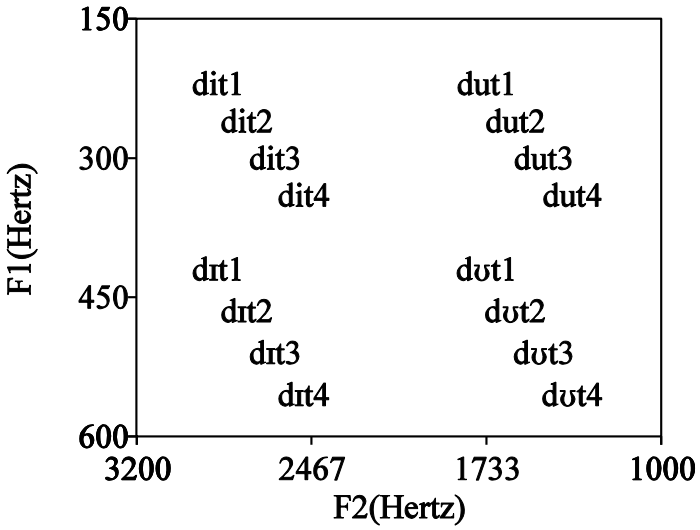

Figure 5.2. Stimuli.

This figure shows a scatter plot of the first (F1) and second (F2) formant frequencies in Hertz of all 4 tokens used per word. /dit1/ and /dut1/ for example have the same F1 but differ in F2, whereas /dit1/ and /dIt1/ have the same F2 but differ in F1. /dit1/ and /dUt1/ neither overlap in F1 nor in F2.

Table 5.1. Testing scheme.

Bird / Group	S+	S-	Probes
729 / 1	/dit/1	/dUt/1	/dIt/1 and /dut/1
728 / 2	/dUt/2	/dit/2	/dIt/2 and /dut/2
750 / 3	/dit/3	/dUt/3	/dIt/3 and /dut/3
763 / 4	/dUt/4	/dit/4	/dIt/4 and /dut/4
734 / 5	/dIt/1	/dut/1	/dit/1 and /dUt/1
731 / 6	/dut/2	/dIt/2	/dit/2 and /dUt/2
758 / 7	/dIt/3	/dut/3	/dit/3 and /dUt/3
741 / 8	/dut/4	/dIt/4	/dit/4 and /dUt/4

This table shows which tokens of which stimuli were presented as either positive or negative stimulus and probes to individual birds and groups of human participants. S+, positive stimulus; S-, negative stimulus.

Human testing

Testing took place in a quiet room using a PC and a custom-written script in the software E-Prime version 2.0. Stimuli were presented via headphones (Sennheiser HD595). Participants learned to discriminate between two of the syllables, following the same Go/NoGo procedure applied to the birds (Table 5.1). Subjects were randomly allocated to the different test groups (1 to 8) with five persons per group and instructed to follow the instructions displayed in Dutch on the computer screen until a note appeared which announced the end of the experiment. Furthermore it was pointed out that during the experiment something might change, but that they were expected to just continue with the procedure. The Go/NoGo paradigm was not explained beforehand so that the human subjects, just like the birds, had to figure out the correct procedure completely by themselves. The experiment started with the screen displaying the instruction: "Press 'Q' to start the trial". After a subject pressed the button 'Q' either the positive or negative stimulus was played back, followed by the instruction: "Press 'P' after the positive stimulus". A two second interval followed in which the subjects had time to press 'P'. Pressing 'P' after the positive stimulus resulted in the presentation of a happy smiley accompanied by a rewarding 'ding' sound. Not pressing 'P' during these two seconds resulted in the presentation of a sad smiley accompanied by a punishing 'attack' sound. After playback of the negative stimulus pressing 'P' resulted in the presentation

of the sad smiley accompanied by the 'attack' sound whereas not pressing 'P' resulted in the presentation of the happy smiley and the 'ding' sound. After this cycle had been completed a new cycle started, again with the instruction: "Press 'Q' to the start the trial" until a total of 10 positive and 10 negative stimulus presentations had taken place. The order of stimulus presentations was random with no more than three positive or negative stimulus playbacks in a row. If a subject had at least 14 correct responses within the first 20 trials (70%) he or she automatically continued to the actual testing phase which was announced by the note: "You are entering the actual testing procedure now". If a subject did not reach the 70% correct responses criterion he or she automatically underwent another training round which was indicated by the sentence: "Your correct score is too low. You will enter another training round.". If a subject still did not achieve 70% correct responses in this second training he or she did not continue to the testing phase and the computer program was terminated with the note: "This is the end of the test. Thank you very much for your participation.". During the testing phase two probe sounds were presented next to the positive and negative stimulus. Each stimulus was presented 16 times in a random order with no more than three consecutive presentations of the same stimulus, resulting in a total of 64 trials. Contrary to the training phase no feedback at all was provided in the testing phase. After the 64 trials a note appeared announcing the end of the experiment and thanking the participants for their participation. The responses to all sounds were automatically saved in E-Prime. The results of one participant of group 7 were not included in the analysis since this person reported to have forgotten which the original positive and negative stimulus was during the testing phase resulting in an 'inverse response', i.e. during testing this person responded to the negative but not to the positive stimulus. Informed consent was obtained from the human participants after the nature of the experiment had been explained.

Results

In the initial training procedure the birds learned to discriminate between the two syllables that differed in all of the formant frequencies (Fig. 5.2 and Table 5.1) after 2143 trials on average (2143.13 ± 257.88 s.e.m., n = 8) following the criterion described earlier (Ohms *et al.* 2010a).

After this initial discrimination stage the two non-reinforced probe sounds were introduced in 20% of the stimulus presentations. The response pattern of the birds to these probe sounds compared to the training stimuli is given in figure 5.3 a,b. Recall that Canadian-English infants used only F1 differences, but not F2 or F3, to distinguish between the vowels of d-vowel-t syllables (Curtin *et al.* 2009). The results of the present study are reversed for zebra finches: they utilized F2 and F3 differences to a greater extent than F1 differences because they categorized stimuli primarily based on differences in F2 and F3 (Fig. 5.3 a,b) by responding to probe sounds that had the same F2 and F3 frequencies as the positive stimulus while ignoring probe sounds with the same F2 and F3 frequencies as the negative stimulus. In other words birds did not weight F1 differences between sounds as strong as F2 and F3 differences as they responded similarly to stimuli and probe sounds which differed in F1. Thus, if a bird was trained to respond to e.g. /dit/ it also responded to /dIt/, whereas if it was trained to respond to /dUt/ it also responded to /dut/. Therefore zebra finches indeed seem to weight higher frequencies, i. e. those for which their auditory system is more sensitive, stronger.

Surprisingly, the results of the human subjects (n = 39, average 24.28 years, ranging from 19 to 34 years) are highly similar compared to the results of the zebra finches (Fig. 5.3) and therefore opposite to the classification pattern of the babies found in earlier studies (Lacerda 1993, 1994; Curtin *et al.* 2009).

Repeated measures ANOVA revealed that the human subjects responded significantly slower to probe sounds compared to training stimuli ($n=39$, $F=7.519$, $p< 0.01$) indicating that they did perceive a difference between the sounds but nevertheless treated them as tokens of the same category. For the birds on the other hand no significant difference between reaction times was detected ($n=8$, $F=0.764$, $p=0,383$) although it is highly likely that they also perceived a difference between training and test stimuli since they responded significantly less to the probe sound that they otherwise treated like the positive stimulus (Fig 5.3 a,b).

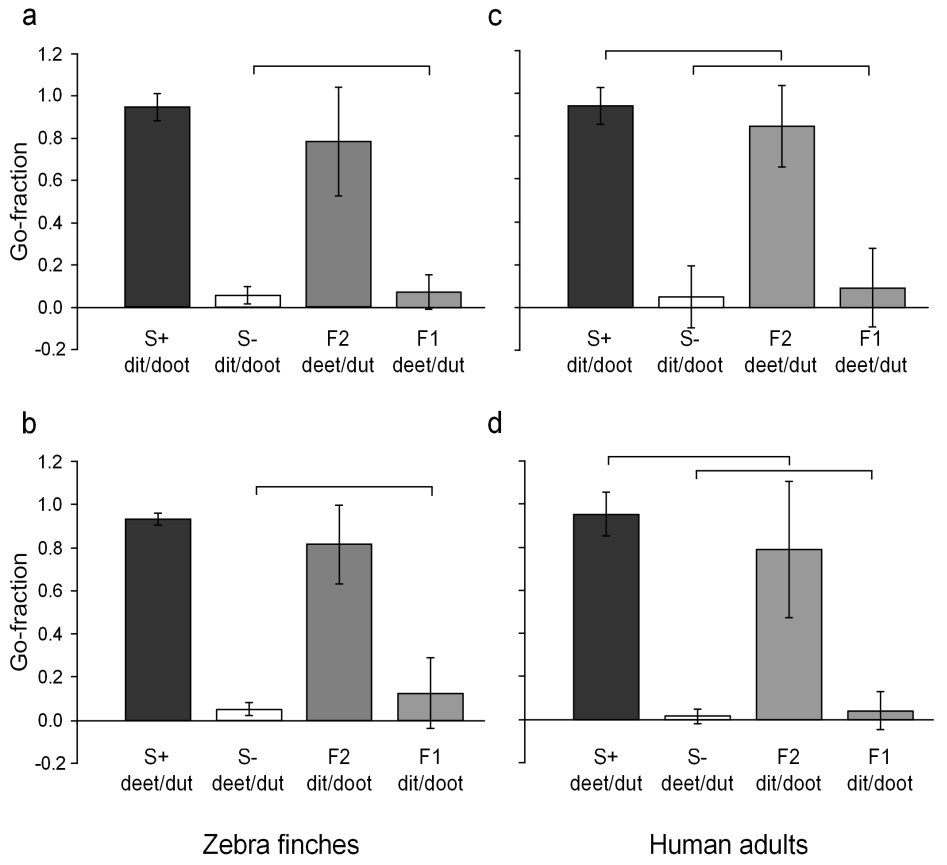

Figure 5.3. Categorization patterns of training stimuli and probe sounds of both zebra finches and Dutch adults.

This figure shows the average proportions including standard deviation of go-responses of birds and humans to training and test stimuli. Every bird got between 50 and 100 probe sound presentations, whereas every human subject got 16 presentations per probe. Horizontal brackets indicate which go-responses did not differ significantly from each other ($p<0.05$) analyzed with a simultaneous testing procedure based on G-tests of independence (Sokal & Rolf 1995). **(a)** and **(c)**, Go-responses of zebra finches (n = 4) and humans (n = 19) respectively that were first trained to discriminate /dIt/ and /dut/ and afterwards got /dit/ and /dUt/ as probe sounds. F2 beneath the bars indicates the go-response to the probe sound that had the same F2 and F3 frequencies as the positive stimulus but the same F1 frequency as the negative stimulus, whereas F1 indicates the go-response to the probe sound that had the same F1 frequency as the positive stimulus but the same F2 and F3 frequencies as the negative stimulus. **(b)** and **(d)**, show the same information as panels (a) and (c) but for those birds (n = 4) and humans (n = 20) that were trained to discriminate /dit/ and /dUt/ and got /dIt/ and /dut/ as probe sounds. S+, positively reinforced stimulus; S-, negatively reinforced stimulus.

Discussion

The results of our study are striking as they reveal a hitherto undiscovered parallel in speech perception between humans and birds. Up to now differences in acoustic cue weighting strategies in speech perception have been attributed to developmental differences between ages (Curtin *et al.* 2009; Nittrouer 1996; Mayo *et al.* 2003; Mayo & Turk 2004) and linguistic background (Escudero *et al.* 2009; Ylinen *et al.* 2009). We now added a new perspective on cue weighting differences by including a non-related, but highly vocal, species. The discovery that both zebra finches and adult Dutch listeners exhibit the same cue weighting strategy for vowel perception might be explained by the fact that both humans and birds show increased sensitivity in higher frequency regions between approximately 1 and 4 kilohertz, i.e. it might not be attributed to linguistic background at all, given that zebra finches obviously lack comparable experience with the Dutch language.

Why then do Canadian-English infants at 15 months of age as well as Swedish infants between 3 and 12 months exhibit an opposite cue weighting bias? Maybe the reason for that lies in initial difficulties of the auditory system to process noisy sounds or sound components that are spectrally less prominent (Nittrouer & Lowenstein 2009) such as F2 and F3 whereas F1, the most prominent spectral feature of a vowel, dictates categorization in an early stage of vocal learning. For normally raised adult zebra finches, which lack experience with human speech, the sensitivity matches the region with the most prominent frequency range of their natural songs. Whether this sensitivity arises from their exposure to a rich conspecific acoustic environment consisting, like human speech, of complex broad-band, amplitude- and frequency-modulated sounds (Lachlan *et al.* 2010) or whether their sensitivity is independent of such an acoustic experience remains an open question. Whatever the causes, our findings do demonstrate that acoustic cue weighting underlying vowel perception in humans does not need to be a highly derived feature linked to the evolution of speech.

Acknowledgements

We thank Dirk Vet from the Institute of Phonetic Sciences at the University of Amsterdam for helping us writing the E-Prime script to test human listeners and all subjects who volunteered to participate in this study. Gabriël J.L. Beckers from the Max-Planck-Institute for Ornithology helped extracting latencies from the data files of the birds. Funding was provided by the Netherlands Organization for Scientific Research (NWO).

Appendix

Table A 5.1. Formant values of the synthesized stimuli.

	/dit/		/dIt/		/dut/		/dUt/	
	F1	**F2**	**F1**	**F2**	**F1**	**F2**	**F1**	**F2**
T1	220	2862	420	2862	220	1736	420	1736
T2	260	2742	465	2742	260	1616	465	1616
T3	300	2622	510	2622	300	1496	510	1496
T4	340	2502	555	2502	340	1376	555	1376

Table A 5.1 gives the frequency values in Hertz of the first two formants of all synthesized stimuli used in this study. T, token; F1, first formant; F2, second formant.

References

A

Anderson, S. R. (2008). The logical structure of linguistic theory. *Language* 84: 795-814.

Assmann, P. F. and Nearey, T. M. (2008). Identification of frequency-shifted vowels. *Journal of the Acoustical Society of America* 124: 3203-3212.

B

Ballentijn, M. R. and ten Cate, C. (1998). Sound production in the collared dove: a test of the 'whistle' hypothesis. *Journal of Experimental Biology* 201: 1637-1649.

Baptista, L. F. and Schuchmann, K. L. (1990). Song learning in the Anna hummingbird (*Calypte anna*). *Ethology* 84: 15-26.

Beckers, G. J. L., Suthers, R. A. and ten Cate, C. (2003). Pure-tone birdsong by resonance filtering of harmonic overtones. *Proceedings of the National Academy of Sciences USA* 100: 7372-7376.

Beckers, G. J. L., Nelson, B. S. and Suthers, R. A. (2004). Vocal-tract filtering by lingual articulation in a parrot. *Current Biology* 14: 1592-1597.

Boersma, P. (2001). PRAAT, a system for doing phonetics by computer. *Glot International* 5: 341-345.

Bolhuis, J. J., Okanoya, K. and Scharff, C. (2010). Twitter evolution: converging mechanisms in birdsong and human speech. *Nature Reviews Neuroscience* 11: 747-759.

Braaten, R. F. and Reynolds, K. (1999). Auditory preference for conspecific song in isolation-reared zebra finches. *Animal Behaviour* 58: 105-111.

Brenowitz, E. A., Perkel, D. J. and Osterhout, L. (2010). Language and birdsong: introduction to the special issue. *Brain and Language* 115: 1-2.

Burdick, C. K. and Miller, J. D. (1975). Speech perception by the chinchilla: discrimination of sustained /a/ and /i/. *Journal of the Acoustical Society of America* 58: 415-427.

C

Castro, L., Medina, A. and Toro, M. A. (2004). Hominid cultural transmission and the evolution of language. *Biology and Philosophy* 19: 721-737.

Clayton, N. S. (1989). The effects of cross-fostering on selective song learning in estrildid finches. *Behaviour* 109: 163-175.

Clench, M. H. (1978). Tracheal elongation in birds-of-paradise. *Condor* 80: 423-430.

Collins, S. A. (2000). Men's voices and women's choices. *Animal Behaviour* 60: 773-780.

Creelman, C. D. (1957). Case of the unknown talker. *Journal of the Acoustical Society of America* 29: 655.

Curtin, S., Fennell, C. and Escudero, P. (2009). Weighting of vowel cues explains patterns of word-object associative learning. *Developmental Science* 12: 725-731.

D

Daley, M. and Goller, F. (2004). Tracheal length changes during zebra finch song and their possible role in upper vocal tract filtering. *Journal of Neurobiology* 59: 319-330.

Dent, M. L., Brittan-Powell, E. F., Dooling, R. J. and Pierce, A. (1997). Perception of synthetic /ba/-/wa/ speech continuum by budgerigars *(Melopsittacus undulatus). Journal of the Acoustical Society of America* 102: 1891-1897.

Diehl, R. L., Lotto, A. J. and Holt, L. L. (2004). Speech perception. *Annual Review of Psychology* 55: 149-179.

Dooling, R. J. (2004). Audition: can birds hear everything they sing? In *Nature's Music. The Science of Birdsong.* (P. Marler and H. Slabbekoorn, eds), 206-225. San Diego: Elsevier Academic Press.

Dooling, R. J., Okanoya, K. and Brown, S. D. (1989). Speech perception by budgerigars (*Melopsittacus undulatus*): the voiced-voiceless distinction. *Perception and Psychophysics* 46: 65-71.

Dooling, R. J. and Brown, S. D. (1990). Speech perception by budgerigars (*Melopsittacus undulatus*): spoken vowels. *Perception and Psychophysics* 47: 568-574.

Dooling, R. J. Best, C. T. and Brown, S. D. (1995). Discrimination of synthetic full-formant and sinewave /ra-la/ continua by budgerigars (*Melopsittacus undulatus*) and zebra finches (*Taeniopygia guttata*). *Journal of the Acoustical Society of America* 97: 1839-1846.

Doupe, A. J. and Kuhl, P. K. (1999). Birdsong and human speech: common themes and mechanisms. *Annual Review of Neuroscience* 22: 567-631.

Dunbar, R. I. M. (2003). The origin and subsequent evolution of language. In *Language Evolution* (M. H. Christiansen and S. Kirby, eds), pp. 219-234. Oxford: Oxford University Press.

E

Eriksson, J. L. and Villa, A. E. P. (2006). Learning of auditory equivalence classes for vowels by rats. *Behavioural Processes* 73: 348-359.

Escudero, P., Benders, T. and Lipski, S. C. (2009). Native, nonnative and L2 perceptual cue weighting for Dutch vowels: the case of Dutch, German and Spanish listeners. *Journal of Phonetics* 37: 452-465.

F

Fant, G. (1960). *Acoustic theory of speech production.* The Hague: Mouton.

Fitch, W. T. (1999). Acoustic exaggeration of size in birds via tracheal elongation: comparative and theoretical analyses. *Journal of Zoology* 248: 31-48.

Fitch, W. T. (2000). The evolution of speech: a comparative review. *Trends in Cognitive Sciences* 4: 258-267.

Fitch, W. T. (2010). *The Evolution of Language.* Cambridge: Cambridge University Press.

Fitch, W. T. and Kelley, J.P. (2000). Perception of vocal tract resonances by whooping cranes, *Grus americana. Ethology* 106: 559-574.

Fitch, W. T. and Reby, D. (2001). The descended larynx is not uniquely human. *Proceedings of the Royal Society of London Series B- Biological Sciences* 268: 1669-1675.

Fletcher, N. H. and Tarnopolsky, A. (1999). Acoustics of the avian vocal tract. *Journal of the Acoustical Society of America* 105: 35-49.

G

Gentner, T. Q., Fenn, K. M., Margoliash, D. and Nusbaum, H. C. (2006). Recursive syntactic pattern learning by songbirds. *Nature* 440: 1204-1207.

Ghazanfar, A. A., Turesson, H. K., Maier, J. X., van Dinther, R., Patterson, R. D. & Logothetis, N. K. (2007). Vocal-tract resonances as indexical cues in rhesus monkeys. *Current Biology* 17: 425-430.

Ghazanfar, A. A. and Rendall D. (2008). Evolution of human vocal production. *Current Biology* 18: R457-R460.

Goller, F., and Larsen, O. N. (1997). A new mechanism of sound generation in songbirds. *Proceedings of the National Academy of Sciences USA* 94: 14787–14791.

Goller, F. and Cooper, B. G. (2004). Peripheral motor dynamics of song production in the zebra finch. *Annals of the New York Academy of Sciences* 1016: 130-152.

Goller, F., Mallinckrodt, M. J. and Torti, S. D. (2004). Beak gape dynamics during song in the zebra finch. *Journal of Neurobiology* 59: 289-303.

Greenewalt, C. H. (1968). *Bird song: acoustics and physiology.* Washington: Smithsonian Institution Press.

H

Haesler, S, Rochefort, C., Georgi, B., Licznerski, P., Osten, P. and Scharff C. (2007). Incomplete and inaccurate vocal imitation after knockdown of FoxP2 in songbird basal ganglia nucleus Area X. *PLoS Biology* 5: e321.

Hausberger, M., Black, J. M. and Richard, J-P. (1991). Bill opening and sound spectrum in barnacle goose loud calls: individuals with 'wide mouths' have higher pitched voices. *Animal Behaviour* 42: 319-322.

Hauser, M. D., Chomsky, N. and Fitch, W. T. (2002). The faculty of language: what is it, who has it and how did it evolve? *Science* 298: 1569-1579.

Hauser, M. D. and Fitch, W. T. (2003). What are the uniquely human components of the language faculty? In *Language Evolution* (M. H. Christiansen and S. Kirby, eds), pp. 158-181. Oxford: Oxford University Press.

Heidweiller, J. and Zweers, G. A. (1990). Drinking mechanisms in the zebra finch and the Bengalese finch. *Condor* 92: 1-28.

Hienz, R. D., Sachs, M. B. and Sinnott, J. M. (1981). Discrimination of steady-state vowels by blackbirds and pigeons. *Journal of the Acoustical Society of America* 70: 699-706.

Hienz, R. D. and Brady, J. V. (1988). The acquisition of vowel discriminations by nonhuman primates. *Journal of the Acoustical Society of America.* 84: 186-194.

Hienz, R. D., Aleszczyk, C. M. and May, B. J. (1996). Vowel discrimination in cats: acquisition, effects of stimulus level, and performance in noise. *Journal of the Acoustical Society of America* 99: 3656-3668.

Hoese, W. J., Podos, J., Boetticher, N. C. and Nowicki, S. (2000). Vocal tract function in birdsong production: experimental manipulation of beak movements. *Journal of Experimental Biology* 203: 1845-1855.

Homberger, D. G. (1986). The lingual apparatus of the African grey parrot, *Psittacus erithacus* Linne (*Aves: Psittacidae*): description and theoretical mechanical analysis. *Ornithological Monographs* 39: iii-xi, 1-233.

J

Janik, V. M. and Slater, P. J. B. (1997). Vocal learning in mammals. *Advances in the Study of Behaviour* 26: 59-99.

Jarvis, E. D. (2004). Learned birdsong and the neurobiology of human language. *Annals of the New York Academy of Sciences* 1016: 749-777.

Jones, E., Oliphant, T., Peterson, P. *et al.* (2001). Open source scientific tools for python. Available: http://www.scipy.org/.

K

Kluender, K. R., Diehl, R. L. and Killeen, P. R. (1987). Japanese quail can learn phonetic categories. *Science* 237: 1195-1197.

Kuhl, P. K. (1981). Discrimination of speech by nonhuman animals: basic auditory sensitivities conducive to the perception of speech-sound categories. *Journal of the Acoustical Society of America* 70: 340-349.

Kuhl, P. K. and Miller, J. D. (1975). Speech perception by the chinchilla: voiced-voiceless distinction in alveolar plosive consonants. *Science* 190: 69-72.

Kuhl, P. K. and Miller, J. D. (1978). Speech perception by the chinchilla: identification functions for synthetic VOT stimuli. *Journal of the Acoustical Society of America* 63: 905-917.

Kuhl, P. K. and Padden, D. M. (1982). Enhanced discriminability at the phonetic boundaries for the voicing feature in macaques. *Perception and Psychophysics* 32: 542-550.

L

Lacerda, F. (1993). Sonority contrasts dominate young infants' vowel perception. *Journal of the Acoustical Society of America* 93: 2372.

Lacerda, F. (1994). The asymmetric structure of the infant's perceptual vowel space. *Journal of the Acoustical Society of America* 95: 3016.

Lachlan, R. F., Peters, S., Verhagen, S. L. and ten Cate, C. (2010). Are there species-universal categories in bird song phonology and syntax? A comparative study of chaffinches (*Fringilla coelebs),* zebra finches (*Taeniopygia guttata*) and swamp sparrows (*Melospiza georgiana*). *Journal of Comparative Psychology* 124: 92-108.

Ladefoged, P. (2006). *A Course in Phonetics.* Boston: Thomson Wadsworth.

Lai, C. S. L., Fisher, S. E., Hurst, J. A., Vargha-Khadem, F. and Monaco, A. P. (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* 413: 519-523.

Larsen, O. N. and Goller, F. (2002). Direct observation of syringeal muscle function in songbirds and a parrot. *Journal of Experimental Biology* 205: 25-35.

Liberman, A. M. (1982). On finding that speech is special. *American Psychologist* 37: 148-167.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P. and Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Revview* 74: 431-461.

Liberman, A. M. and Mattingly, I. G. (1985). The motor theory of speech revised. *Cognition* 21: 1-36.

Liberman, A. M. and Whalen, D. H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences* 4: 187-196.

Lieberman, P., Klatt, D. H. and Wilson, W. H. (1969). Vocal tract limitations on the vowel repertoires of rhesus monkey and other nonhuman primates. *Science* 164: 1185-1187.

Lieberman, P. (1975). *On the origins of language. An introduction to the evolution of human speech.* New York: Macmillan.

Lieberman, P. (1984). *The biology and evolution of language.* Cambridge: Harvard University Press.

M

Macmillan, N. A. and Creelman, C. D. (2005). *Detection Theory. A User's Guide.* Mahwah: Lawrence Erlbaum Associates.

Magnuson, J. S. and Nusbaum, H. C. (2007). Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *Journal of Experimental Psychology- Human Perception and Performance* 35: 391-409.

Marler, P. (1976). An ethological theory of the origin of vocal learning. *Annals of the New York Academy of Sciences* 280: 386-395.

Martella, M. B. and Bucher, E. H. (1990). Vocalizations of the monk parakeet. *Bird Behaviour* 8: 101-110.

Mayo, C., Scobbie, J. M., Hewlett, N. and Waters, D. (2003). The influence of phonemic awareness development on acoustic cue weighting strategies in children's speech perception. *Journal of Speech and Hearing Research* 46: 1184-1196.

Mayo, C. and Turk, A. (2004). Adult-child differences in acoustic cue weighting are influenced by segmental context: Children are not always perceptually biased towards transitions. *Journal of the Acoustical Society of America* 115: 3184-3194.

Mullennix, J. W., Pisoni, D. B. and Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America* 85: 365-378.

N

Nearey, T. M. (1989). Static, dynamic, and relational properties in vowel perception. *Journal of the Acoustical Society of America* 85: 2088-2133.

Nelson, B. S., Beckers, G. J. L. and Suthers, R. A. (2005). Vocal tract filtering and sound radiation in a songbird. *Journal of Experimental Biology* 208: 297-308.

Nishimura, T., Mikami, A., Suzuki, J. and Matsuzawa, T. (2006). Descent of the hyoid in chimpanzees: evolution of face flattening and speech. *Journal of Human Evolution* 51: 244-254.

Nittrouer, S. (1996). The relation between speech perception and phonemic awareness: evidence from low-SES children and children with chronic OM. *Journal of Speech and Hearing Research* 39: 1059-1070.

Nittrouer, S. and Lowenstein, J. H. (2009). Does harmonicity explain children's cue weighting of fricative-vowel syllables? *Journal of the Acoustical Society of America* 125: 1679-1692.

Nottebohm, F. (1976). Phonation in the orange-winged Amazon parrot, *Amazona amazonica*. *Journal of Comparative Physiology* 108: 157–170.

Nowicki, S. (1987). Vocal tract resonances in oscine bird sound production: evidence from birdsongs in a helium atmosphere. *Nature* 325: 53-55.

Nowicki, S. and Capranica, R. R. (1986). Bilateral syringeal interaction in vocal production of an oscine bird sound. *Science* 231: 1297-1299.

O

Ohms, V. R., Gill, A., van Heijningen, C. A. A., Beckers, G. J. L. and ten Cate, C. (2010). Zebra finches exhibit speaker-independent phonetic perception of human speech. *Proceedings of the Royal Society of London Series B- Biological Sciences* 277: 1003-1009.

Ohms, V. R., Snelderwaard, P. C., ten Cate, C. and Beckers, G. J. L. (2010). Vocal tract articulation in zebra finches. *PLoS ONE* 5: e11923.

P

Patterson, D. K. and Pepperberg, I. M. (1994). A comparative study of human and parrot phonation: acoustic and articulatory correlates of vowels. *Journal of the Acoustical Society of America* 96: 634-648.

Pepperberg, I. M. (2010). Vocal learning in grey parrots: a brief review of perception, production and cross-species comparisons. *Brain and Language* 115: 81-91.

Pepperberg, I. M., Howell, K. S., Banta, P. A., Patterson, D. K. and Meister, M. (1998). Measurement of grey parrot (*Psittacus erithacus*) trachea via magnetic resonance imaging, dissection and electron beam computed tomography. *Journal of Morphology* 238: 81-91.

Peterson, G. E. and Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America* 24: 175-184.

Pinker, S. and Jackendoff, R. (2005). The faculty of language: what's special about it? *Cognition* 95: 201-236.

Pinker, S. (2010). The cognitive niche: coevolution of intelligence, sociality and language. *Proceedings of the National Academy of Sciences USA* 107: 8993-8999.

Plummer, E. M. and Goller, F. (2008). Singing with reduced air sac volume causes uniform decrease in airflow and sound amplitude in the zebra finch. *Journal of Experimental Biology* 211: 66-78.

Podos, J., Sherer, J. K., Peters, S. and Nowicki, S. (1995). Ontogeny of vocal tract movements during song production in song sparrows. *Animal Behaviour* 50: 1287-1296.

Podos, J., Southall, J. A. and Rossi-Santos, M. R. (2004). Vocal mechanics in Darwin's finches: Correlation of beak gape and song frequency. *Journal of Experimental Biology* 207: 607-619.

Polka, L. and Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology- Human Perception and Performance* 20: 421-435.

Poole, J. H., Tyack, P. L., Stoeger-Horwath, A. S. and Watwood, S. (2005). Elephants are capable of vocal learning. *Nature* 434: 455-456.

R

Reby, D., McComb, K., Cargnelutti, B., Darwin, C., Fitch, W. T. & Clutton-Brock, T. (2005). Red deer stags use formants as assessment cues during intrasexual agonistic interactions. *Proceedings of the Royal Society of London Series B- Biological Sciences* 272: 941-947.

Riede, T., Beckers, G. J. L., Blevins, W. and Suthers, R. A. (2004). Inflation of the esophagus and vocal tract filtering in ring doves. *Journal of Experimental Biology* 207: 4025-4036.

Riede, T., Suthers, R. A., Fletcher, N. H and Blevins, W. E. (2006). Songbirds tune their vocal tract to the fundamental frequency of their song. *Proceedings of the National Academy of Sciences USA* 103: 5543-5548.

Riede, T. and Suthers, R. A. (2009). Vocal tract motor patterns and resonance during constant frequency song: the white-throated sparrow. *Journal of Comparative Physiology A- Neuroethology, Sensory, Neural and Behavioural Physiology* 195: 183-192.

S

Snelderwaard, P. C., de Groot, J. H. and Deban, S. M. (2002). Digital video combined with conventional radiography creates an excellent high-speed X-ray video system. *Journal of Biomechanics* 35: 1007-1009.

Smith, D. R. R. and Patterson, R. D. (2005). The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age. *Journal of the Acoustical Society of America* 118: 3177-3186.

Sokal, R. R. and Rolf, F. J. (1995). *Biometry: The Principles and Practice of Statistics in biological Research.* New York: W. H. Freeman and Company.

Suthers, R. A. (1990). Contributions to birdsong from the left and right sides of the intact syrinx. *Nature* 347: 473-477.

Suthers, R. A. (1999). Peripheral control and lateralization of birdsong. *Journal of Neurobiology* 33: 632-652.

Suthers, R. A., Goller, F. and Hartley, R. S. (1994). Motor dynamics of song production by mimic thrushes. *Journal of Neurobiology* 25: 917-936.

Suthers, R. A., Vallet, E., Tanvez, A. and Kreutzer, M. (2004). Bilateral song production in domestic canaries. *Journal of Neurobiology* 60: 381-393.

Suthers, R. A. and Zollinger, S. A. (2004). Producing song- The vocal apparatus. *Annals of the New York Academy of Sciences* 1016: 109-129.

T

Titze, I. R. (2000). *Principles of Voice Production.* Iowa City: National Center for Voice and Speech.

Trout, J. D. (2003). Biological specializations for speech: what can the animals tell us? *Current Direction in Psychological Science* 12: 155-159.

V

van Heijningen, C. A. A., de Visser, J., Zuidema, W. and ten Cate, C. (2009). Simple rules can explain discrimination of putative recursive syntactic structures by a songbird species. *Proceedings of the National Academy of Sciences USA* 106: 20538-20543.

Verzijden, M. N., Etman, E., van Heijningen, C. A. A., van der Linden, M. and ten Cate, C. (2007). Song discrimination learning in zebra finches induces highly divergent responses to novel songs. *Proceedings of the Royal Society of London Series B- Biological Sciences* 274: 295-301.

W

Warren, D. K., Patterson, D. K. and Pepperberg, I. M. (1996). Mechanisms of American English vowel production in a grey parrot (*Psittacus erithacus*). *The Auk* 113: 41-58.

Westneat, M.W., Long, J. H., Hoese, W. and Nowicki, S. (1993). Kinematics of birdsong: functional correlation of cranial movements and acoustic features in sparrows. *Journal of Experimental Biology* 182: 147-171.

Whaling, C. S., Solis, M. M., Doupe, A. J., Soha, J. A. and Marler, P. (1997). Acoustic and neural bases for innate recognition of song. *Proceedings of the National Academy of Sciences USA* 94: 12694-12698.

White, S. A. (2001). Learning to communicate. *Current Opinion in Neurobiology* 11: 510-520.

Williams, H. (2001). Choreography of song, dance and beak movements in the zebra finch *(Taeniopygia guttata). Journal of Experimental Biology* 204: 3497-3506.

Y

Ylinen, S., Uther, M., Latvala, A., Vepsäläinen, S., Iverson, P., Akahane-Yamada, R. and Näätänen, R. (2009). Training the brain to weight speech cues differently: a study of Finnish second-language users of English. *Journal of Cognitive Neuroscience* 22: 1319-1332.

Yule, G. (2006). *The study of language.* Cambridge: Cambridge University Press.

Z

Zollinger, S. A. and Suthers, R. A. (2004). Motor mechanisms of a vocal mimic: implications for birdsong production. *Proceedings of the Royal Society of London Series B- Biological Sciences* 271: 483-491.

Zollinger, S. A., Riede, T. and Suthers, R. A. (2008). Two-voice complexity from a single side of the syrinx in the northern mockingbird *Mimus polyglottos* vocalizations. *Journal of Experimental Biology* 211: 1978-1991.

Nederlandse samenvatting

Spraak bij verschillende soorten:

Mechanistische grondlagen van vocale productie en perceptie

Dit is een vertaling van hoofdstuk 1. De literatuurverwijzingen zijn weggelaten om de leesbaarheid te verbeteren. De figuur is terug te vinden in hoofdstuk 1.

Algemene inleiding

Menselijke taal is één van de meest complexe gedragingen die wij kennen. Zij maakt het mogelijk oneindig veel ideeën te ontwikkelen en te communiceren. Hoewel het belang van het menselijke vermogen tot taal evident is, is tot dusver onduidelijk, en is er veel discussie over, hoe taal evolutionair ontstaan is en welke specifieke eigenschappen menselijke taal uniek maken.

Spraak is het primaire fysieke fenomeen waarmee taal wordt overgedragen. Een beperkt aantal betekenisloze geluiden wordt gecombineerd om een potentieel oneindige set van betekenisvolle woorden te vormen. De onderliggende productie- en perceptiemechanismen kunnen met behulp van akoestische, fysiologische, anatomische en neurobiologische methoden worden bestudeerd. Helaas kunnen dit soort studies ons echter weinig leren over de evolutie van spraak. Ook fossiele vondsten, voor zover voorhanden, leveren te weinig informatie om de evolutie van spraak in kaart te kunnen brengen. Het bestuderen van vocale communicatie bij dieren maakt het echter wel mogelijk om mechanismen op te sporen die homoloog of door convergentie zijn ontstaan, en kan daarom helpen onderliggende evolutionaire processen te identificeren. Deze vergelijkende benadering is één van de belangrijkste methoden van een nog jong wetenschapsgebied dat biolinguïstiek wordt genoemd.

Bevindingen van de vergelijkende benadering

In de laatste jaren zijn in een groeiend aantal studies mogelijke overeenkomsten tussen menselijke en dierlijke vocale communicatie onderzocht. Eén van de belangrijkste eigenschappen van spraak is dat het geleerd wordt door imitatie, iets wat zeldzaam is in de ontwikkeling van vocalisatie bij andere dieren. Binnen de zoogdieren komt vocaal leren, behalve bij de mens, slechts in een beperkt aantal niet-verwante groepen voor, waaronder zeehonden, walvissen, vleermuizen en olifanten. Opmerkelijk is dat vocaal leren bij onze naaste verwanten, mensapen en andere primaten, niet voorkomt.

Behalve bij zoogdieren komt vocaal leren ook voor in drie vogelgroepen, namelijk de zangvogels, papegaaien en kolibries. Deze en andere parallellen tussen menselijke spraak en vogelvocalisatie hebben ertoe geleid dat vogelzang als het meest geschikte diermodel wordt beschouwd om onderliggende mechanismen van spraakontwikkeling, -productie en -perceptie te bestuderen.

Zowel mensen als vogels vertonen vroeg in het leven een sensitieve fase waar vocaal leren wordt gefaciliteerd. Auditieve feedback door het horen van eigen geproduceerde

vocalisaties en vocalisaties van anderen is hierbij cruciaal. Er wordt ook gedacht dat aangeboren predisposities in mensen en vogels bepalen welke geluiden worden geleerd. Sinds kort is het dankzij de ontwikkeling van nieuwe moleculaire methoden ook mogelijk om de genetische basis van vocaal gedrag te onderzoeken. Er is een genmutatie gevonden die een spraakstoornis bij mensen veroorzaakt en er bij zebravinken toe leidt dat ze de zang van hun tutor onnauwkeurig kopiëren.

De hierboven beschreven parallellen hebben vooral betrekking op vocale ontwikkeling en leren. Ook op het gebied van vocale productie en perceptie lijken echter overeenkomsten tussen vogels en mensen te bestaan. Vogelzang en spraak worden beide door een geluidsbron geproduceerd en daarna in het spraakkanaal gefilterd, hoewel over de details hiervan bij vogels minder bekend is. Wat betreft vocale perceptie is nog onduidelijk of er speciale perceptuele capaciteiten nodig zijn voor het waarnemen en onderscheiden van spraak, en of zangvogels deze capaciteiten ook hebben. In dit proefschrift onderzoek ik zowel vocale productie- als perceptiemechanismen in (zang)vogels en vergelijk ik deze met die van mensen om erachter te komen welke overeenkomsten en verschillen er zijn.

Formanten en hun relevantie in vocale communicatie

Menselijke spraak wordt gekarakteriseerd door een breed spectrum van verschillende frequenties. Stemhebbende geluiden worden in de larynx geproduceerd door het vibreren van de stembanden. Daardoor ontstaat een fundamentele frequentie met harmonische boventonen. Dit geluid wordt in het spraakkanaal, bestaande uit keel-, mond- en neusholtes, gefilterd. Afhankelijk van de positie van de articulatoren, zoals tong en lippen, worden hierbij sommige frequenties versterkt. Deze frequenties worden 'spraakkanaalresonanties' ofwel 'formanten' genoemd. Formanten zijn in een spectrogram als zwarte banden herkenbaar (Fig. 1.1). Formantpatronen worden onafhankelijk van de stembanden gevormd, en spelen vooral bij de productie en perceptie van klinkers een belangrijke rol. Het verschil tussen 'wit' en 'wet', bijvoorbeeld, is gebaseerd op verschillende formantwaarden, vooral met betrekking tot de twee laagste formanten. Deze worden veroorzaakt door resonanties van de keel- (F1) en mondholte (F2).

Tijdens de menselijke ontogenese daalt de larynx, waardoor de tong zowel horizontaal als verticaal kan bewegen. Dit maakt het mogelijk om de geometrie van het spraakkanaal op complexe manieren te moduleren, waardoor resonantiefrequenties veranderen en complexe formantenpatronen kunnen ontstaan (Fig. 1.1).

In het algemeen wordt aangenomen dat het dalen van de larynx en het verlies van luchtzakken rond de larynx een noodzakelijke voorwaarde is geweest voor de evolutie

van spraak. Het is echter niet waarschijnlijk dat spraakevolutie een veroorzakende kracht is geweest bij het ontstaan van deze anatomische configuratie, gezien het feit dat ook andere diersoorten een gedaalde larynx vertonen. Zo verlagen edelherten en damherten hun larynx tijdens hun burlen, en hebben sommige vogelsoorten verlengde luchtpijpen. In beide gevallen leidt dit tot lagere resonantiefrequenties van het spraakkanaal, waardoor deze individuen groter klinken. Het is dus waarschijnlijk dat het zakken van de larynx door seksuele selectie werd veroorzaakt, en daarmee een preadaptatie geworden is voor de evolutie van spraak. Bovendien hebben ook chimpansees een verlaagde larynx en wordt soms gespeculeerd dat het vervlakken van het gezicht belangrijker was voor het ontstaan van de karakteristieke configuratie van het spraakkanaal.

Over spraakkanaalresonanties in zangvogels is echter minder bekend. Het "spraakorgaan" van vogels, de *syrinx* genoemd, is veel gecompliceerder dan de menselijke larynx. Zangvogels bezitten twee paar vibrerende membranen die aan het eind van elke bronchus zitten, en die betrokken zijn bij geluidsproductie. Deze twee stembandsets kunnen onafhankelijk van elkaar worden gecontroleerd, waardoor de vogel met twee stemmen tegelijk kan zingen of kan afwisselen. Vanwege deze complexiteit op stemniveau verwachtte men aanvankelijk dat alle complexiteit in vogelzang door de geluidsbron wordt veroorzaakt en dat filtermechanismen zoals in menselijke spraak geen rol spelen. Recent onderzoek laat echter zien dat zang- en spraakproductie meer gemeen hebben dan oorspronkelijk werd aangenomen. Eén van de doelen van dit proefschrift is om een beter begrip te verkrijgen van potentiële articulatoren in vogelzang en hun invloed op het geproduceerde geluid.

Het tweede doel is om een beter begrip te verkrijgen van hoe formanten door vogels waargenomen worden. Mensen zijn erg gevoelig voor formantenpatronen en spraakvariatie. Het is echter niet duidelijk of deze sensitiviteit in samenhang met spraakproductie is geëvolueerd, en dus uniek is voor mensen, of dat deze berust op algemene eigenschappen van auditieve systemen. Sommige perceptuele fenomenen waarvan men dacht dat ze uniek waren voor de mens, zoals categorische perceptie, zijn ook bij zoogdieren en vogels gevonden. Een ander belangrijk kenmerk van spraak is dat wij woorden kunnen herkennen ondanks het feit dat deze akoestisch sterk kunnen verschillen wanneer ze door verschillende sprekers uitgesproken worden. Dit fenomeen wordt sprekernormalisatie genoemd. Het is onbekend of dit een fenomeen is dat alleen bij mensen voorkomt, en dus waarschijnlijk in samenhang met spraak is ontstaan, of dat het veroorzaakt wordt door algemene eigenschappen van het auditieve systeem.

Dit proefschrift

In dit proefschrift worden vier experimenten beschreven, waarvan twee over productie- en twee over perceptiemechanismen van formanten bij vogels gaan. Ik heb vooral zebravinken voor mijn experimenten gebruikt omdat deze het meest bestudeerde modelsysteem zijn voor vocaal leren, en er desondanks weinig bekend is over geluidsproductie en -perceptie in deze soort. Daarnaast heb ik ook geluidsproductiemechanismen bij monniksparkieten onderzocht, omdat er indicaties voor belangrijke verschillen tussen zangvogels en papegaaien zijn.

In **hoofdstuk 2** beschrijf ik een experiment dat is uitgevoerd om potentiële articulatoren in zebravinken te identificeren en hun betrokkenheid bij zangproductie te evalueren. Eerst hebben wij röntgenfilms van zingende zebravinken gemaakt, waarvan de analyse heeft laten zien dat vooral snavelopening en het vergroten van de orofaryngeale-esofageale holte (OEC) als articulatoren dienen. Dit resultaat is in overeenstemming met eerdere studies die positieve correlaties tussen snavelopening en frequentiepatronen in meerdere zangvogelsoorten, waaronder zebravinken, hebben aangetoond. Interessanter is de observatie dat zebravinken hun OEC sterk uitbreiden tijdens de zang. Dit is eerder in de rode kardinaal en de witkeelgors aangetoond. Deze twee soorten produceren echter allebei een redelijk eenvoudig, tonaal gezang met weinig energie in harmonische boventonen, en uitbreiding van de OEC bleek bij deze soorten te corresponderen met variatie in de fundamentele frequentie. Zebravinken produceren daarentegen een redelijk ingewikkelde zang met veel verschillende elementtypes. De meeste van deze elementen hebben een breed frequentiespectrum en gevarieerde amplitudepatronen. De complexiteit van zebravinkenzang maakt het lastig om duidelijke verbanden tussen articulatorenconfiguratie en geluidspatronen op te sporen. Desalniettemin hebben wij in het tweede deel van de studie, waarin we snavelopening en OEC variatie experimenteel hebben gemanipuleerd, gevonden dat de piekfrequentie zakt tijdens de OEC vergroting terwijl amplitude toeneemt, vooral tussen 1.5 en 4.5 kHz. Snavelopening blijkt frequenties rond de 5 kHz en daarboven te versterken. Deze resultaten tonen aan dat modulatie van het spraakkanaal, vooral door snavelopening en de uitbreiding van de OEC, een belangrijke rol speelt in het dynamisch filteren van zebravinkenzang, wat in ingewikkelde elementtypes met formantachtige patronen resulteert.

Papegaaien vormen een andere vogelgroep die complexe geluiden over een breed frequentiespectrum produceert; deze vogels staan bekend om hun vermogen tot het imiteren van menselijke spraak. In tegenstelling tot zangvogels hebben papegaaien een tong met veel intrinsieke spieren en een vlezig, flexibel oppervlak, die lijkt op de

menselijke tong. Om deze reden gaat men ervan uit dat de tong een veel belangrijkere rol speelt bij de geluidsproductie van papegaaien dan bij die van zangvogels. Observaties van een spraakimiterende papegaai en experimentele manipulatie van tongpositie bij monniksparkieten ondersteunen deze stelling. Tot nu toe waren echter nog geen observaties van tongbewegingen in natuurlijk communicerende papegaaien gedaan, noch was bekend welke andere articulatoren bij geluidsproductie van papegaaien betrokken zijn.

In **hoofdstuk 3** heb ik dit onderzocht door röntgenfilms van natuurlijk communicerende papegaaien te analyseren. Aan de hand van de video's konden wij drie typen articulatorische bewegingen identificeren: snavelopening, veranderingen in de tonghoogte en samentrekking van de luchtpijp. Hoewel eerdere studies het belang van de tong al hebben aangetoond, zijn in deze studies vooral effecten in de horizontale dimensie van de tongpositie gevonden terwijl de parkieten in onze studie vooral tonghoogte veranderden. Van de negen bekende vocalisatietypes die volwassen monniksparkieten produceren, produceerden onze vogels in het laboratorium er maar drie. Daarom is het mogelijk dat de horizontale tongpositie wel een belangrijke rol speelt in de vocalisaties die wij niet konden opnemen. Niettemin blijken veranderingen in de verticale tongdimensie van begroetingsroepen, die wij wel hebben opgenomen en die formantachtige patronen laten zien, belangrijker te zijn dan veranderingen in de horizontale dimensie. Interessant genoeg hebben wij ook bewijzen gevonden voor een samentrekking van de luchtpijp, terwijl een eerdere studie in zebravinken geconcludeerd heeft dat veranderingen in de lengte van de luchtpijp te klein zijn om invloed te hebben op het geluid in deze soort. Verder vonden we positieve correlaties tussen amplitude in begroetingsroepen en schettergeluiden voor snavelopening, tonghoogte en luchtpijpcontractie in sommige van de parkieten. Omdat modulaties in de fundamentele frequentie (F0) redelijk snel zijn in monniksparkietroepen, maar articulatorische bewegingen relatief langzaam, is het waarschijnlijk dat veranderingen in F0 door de geluidsbron worden veroorzaakt. Formantpatronen zoals in begroetingsroepen ontstaan waarschijnlijk wel door de spraakkanaalfilter en worden in dat geval gemoduleerd door de werking van de articulatoren. Helaas was het niet mogelijk om duidelijke relaties tussen formantpatronen en articulatie te bepalen omdat de precieze eigenschappen van de geluidsbron grotendeels onbekend zijn. Derhalve moeten toekomstige studies de precieze aard van deze relaties bepalen, en zijn er meer gegevens nodig over de anatomie en fysiologie van het vocale systeem om een betrouwbaar model van geluidsproductie in deze vogels op te stellen.

In de tweede helft van dit proefschrift wordt formantperceptie bij vogels vergeleken met die bij mensen. Hoewel er verschillen zijn in vocale communicatie bij mensen, zangvogels en papegaaien, maken alle drie groepen gebruik van actieve geluidsfiltermechanismen waardoor ze veel verschillende geluiden kunnen produceren. Dit leidt tot de vraag of de onderliggende mechanismen van formantperceptie ook vergelijkbaar zijn tussen mensen en vogels. Als dat het geval is, dan is het niet noodzakelijk om ervan uit te gaan dat speciale mechanismen voor formantperceptie in mensen zijn geëvolueerd door coëvolutie van spraakproductie en -perceptie. In plaats daarvan zouden algemene verwerkingsmechanismen van het auditieve systeem voldoende zijn om menselijke spraakgeluiden te onderscheiden.

Hoofdstuk 4 beschrijft een studie naar sprekernormalisatie bij zebravinken. Hiervoor zijn natuurlijk geproduceerde menselijke woorden gebruikt, ingesproken door jongvolwassen Nederlandstalige sprekers. Eén van de belangrijkste aspecten van menselijke spraak is ons vermogen woorden te kunnen herkennen onafhankelijk van de spreker en ondanks veel variatie tussen sprekers onderling. Taalwetenschappers gaan ervan uit dat dit een gevolg is van het menselijke vermogen voor extrinsieke en intrinsieke sprekernormalisatie. Intrinsieke sprekernormalisatie verklaart waarom geluiden die als dezelfde fonemen worden waargenomen verschillende akoestische realisaties kunnen hebben door ervan uit te gaan dat elk spraakmonster door een normaliseringtransformatie wordt gecategoriseerd. Tegelijkertijd is bekend dat er een sprekereffect bestaat, wat in het begin spraakdiscriminatie tussen sprekers moeilijker maakt. Dit probleem wordt opgelost via een leerproces waarin een referentieframe van verschillende spraakgeluidmonsters ontstaat. Wij hebben middels operante conditionering acht zebravinken getraind om twee woorden, 'wit' en 'wet', te onderscheiden en deze later te generaliseren naar onbekende sprekers van (1) hetzelfde geslacht en (2) het andere geslacht. De twee woorden verschillen voornamelijk in hun formantenpatroon. Alle acht vogels waren in staat de woorden te onderscheiden en te categoriseren onafhankelijk van de sprekers. Onze analyse heeft laten zien dat ze dit onderscheid aan de hand van de formantpatronen hebben gemaakt. Bovendien gebruikten de vogels, net als mensen, hiervoor een combinatie van extrinsieke en intrinsieke sprekernormalisatie. Dit resultaat impliceert dat formantperceptie ofwel wijdverspreid is in het dierenrijk, ofwel convergent is ontstaan in mensen en vogels.

Het laatste hoofdstuk van dit proefschrift, **hoofdstuk 5**, beschrijft een directe vergelijking van akoestische parameterweging in klinkerperceptie tussen zebravinken en Nederlandse volwassenen. In het verleden is aangetoond dat zowel Canadees-Engelse als Zweedse baby's tussen drie en twaalf maanden oud gevoeliger zijn voor lage

frequenties (F1) wanneer ze woorden onderscheiden. Dit is verrassend omdat algemeen aangenomen wordt dat de taalomgeving een grote invloed heeft op klinkerperceptie vanaf een leeftijd van zes maanden. Dit kan ofwel een indicatie voor een algemene menselijke gevoeligheid voor lage frequenties in klinkerperceptie zijn, ofwel te maken hebben met een rijpingsproces van het auditieve systeem. Maar de vraag of deze gevoeligheden verbonden zijn aan spraakperceptie en dus een uniek menselijke eigenschap zijn, of juist een algemeen karakteristiek zijn voor auditorische perceptie, is nog niet onderzocht. In een eerste poging deze vraag te beantwoorden hebben wij zebravinken en Nederlandse volwassene mensen op dezelfde manier getest. Beide werden in een *'Go/NoGo'*-procedure blootgesteld aan twee gesynthetiseerde woorden (dit en doet of dut en diet) die alleen in hun klinkers van elkaar verschilden en moesten leren tussen die twee woorden onderscheid te maken. De klinkers hadden verschillende F1 en F2 frequenties. In de volgende stap hebben wij naast de eerste twee woorden twee testwoorden gepresenteerd. Deze testwoorden werden nooit beloond en de reacties op de testwoorden liet zien hoe mensen en vogels deze hebben waargenomen. Eén testwoord had dezelfde F1 frequentie als het eerste originele woord en de dezelfde F2 frequentie als het tweede originele woord en andersom voor het tweede testwoord. De reacties van vogels en mensen op de testwoorden waren opvallend overeenkomstig: beide vertoonden ze een neiging om categorieën te maken gebaseerd op hoge frequenties (F2). Dit is precies andersom dan wat baby's doen en toont ten eerste aan dat akoestische parameterweging niet uniek is voor mensen en afhankelijk van spraakperceptie en ten tweede dat de maturatie van het auditieve systeem waarschijnlijk een rol zal spelen bij het ontstaan van een dergelijke parameterweging.

Discussie en conclusie

In dit proefschrift heb ik laten zien dat zebravinken en monniksparkieten, evenals mensen, gebruik maken van verschillende articulatoren om het geluid dat in de syrinx wordt geproduceerd te modificeren. Terwijl in zangvogels vooral snavelopening en OEC uitbreiding een belangrijke rol spelen blijken papegaaien voornamelijk de tong te gebruiken om het geluid te filteren. Dit is vergelijkbaar met menselijke spraakproductie en waarschijnlijk één van de redenen waarom papegaaien spraak zo goed kunnen imiteren. Gebaseerd op deze observaties kan geconcludeerd worden dat de geluidsproductiemechanismen in vogels en mensen meer gemeen hebben dan oorspronkelijk werd vermoed, wat convergentie in evolutionaire patronen suggereert. Mogelijk zijn sommige articulatoren die nu worden gebruikt voor geluidsverandering

oorspronkelijk als delen van het voedselverwerkingssysteem ontstaan. In dat geval zijn ecologische adaptaties voor verschillende voedselsoorten de drijvende kracht achter de evolutie van deze structuren geweest, en zijn deze structuren later door het communicatiesysteem gebruikt om vocale variatie te verhogen.

Met betrekking tot spraakperceptie heb ik laten zien dat zebravinken, net als mensen, formantpatronen gebruiken om op elkaar lijkende woorden te onderscheiden en tegelijkertijd extrinsieke en intrinsieke sprekernormalisatie toepassen om woorden onafhankelijk van de spreker te categoriseren. Dit heeft belangrijke implicaties voor de evolutie van formantperceptie. Zoals al eerder werd gespeculeerd lijkt formantperceptie in veel verschillende diersoorten te zijn geëvolueerd om informatie over geslacht, grootte en leeftijd van een individu beschikbaar te maken. Tijdens de evolutie van spraak zou deze capaciteit vervolgens kunnen hebben geleid tot een communicatiesysteem dat gebruik maakt van formanten om informatie te coderen. Het feit dat volwassen mensen en zebravinken dezelfde parameterweging in klinkerperceptie vertonen versterkt deze hypothese. Beide soorten zijn gevoeliger voor F2 frequenties, welke in het meest sensitieve bereik van het auditieve systeem vallen, bij het categoriseren van meerduidige klinkers. Menselijke baby's aan de andere kant blijken gevoeliger te zijn voor frequenties die sterker worden betoond, namelijk F1.

Samenvattend heb ik laten zien dat (zang)vogels in staat zijn tot formantenproductie en -perceptie en dat de onderliggende mechanismen meer overeenkomen in mensen en vogels dan eerder werd aangenomen. Zowel zangvogels als papegaaien kunnen als modelorganismen worden gebruikt om meer over de selectiedrukken, die tot zulk ingewikkelde communicatiesystemen hebben geleid, te weten te komen. Nu sommige van de onderliggende mechanismen voor formantenproductie en -perceptie zijn geïdentificeerd kan in toekomstige studies begonnen worden met het onderzoek naar de rol die formanten in soorteigen vogelvocalisaties spelen, en met het ontwikkelen van gedetailleerde modellen voor geluidsproductie en -perceptie.

Acknowledgements

Praesidium libertatis- fortress of freedom. I cannot imagine any motto that would better describe how I have experienced Leiden University over the past 4 years since for the first time in my life I had the impression of being accepted for who I am. No one within this university and especially within the Behavioural Biology group has ever tried to pigeonhole me, but accepted me being *'op de kast'* and taken it with humor.

Completing a PhD thesis is an enormous undertaking whose success among other things depends on a stimulating and motivating environment. Therefore I would like to thank my past and current colleagues and collaborators from various places around the world who all contributed to my achievement one way or the other: Amy, Ardie, Aukje, Barbara, Carel, Caroline, Erwin, Fabienne, Gabriël, Hans, Henny, Irena, Jelle, Jiani, Katharina, Kenneth, Kori, Machteld, Marie-Jeanne, Niels, Paola, Paula, Peter, Rod, Sita, Wouter and Xiao Jing.

Finding yourself in a new environment whose language you don't speak can be a major challenge but especially thanks to Aukje and Erwin I mastered.

Arike, Karin, Nico and Corine decided to do their internships with me and I hope you learned just as much from me as I did from you.

Caroline and Karin, thanks for being my paranymphs! You almost seem to be more excited about my graduation than I am myself.

Janine, es tat gut zu wissen, dass du nur eine Etage tiefer mit den gleichen, zum Teil frustrierenden, Dingen zu kämpfen hast, die auch mir manch schlaflose Nacht bereitet haben. Manche Dinge kann man einfach am besten verstehen wenn man im gleichen Sprachraum aufgewachsen ist. Dasselbe gilt für Bernadette: auch wenn es letztendlich vier Wochen gedauert hat bis wir herausgefunden haben, dass wir mehr als nur unser Interesse für Badminton teilen, war es doch allzu offensichtlich, dass eine gewisse Verbindung zwischen uns besteht. Dank euch beiden für euren Eifer, Enthusiasmus und kritischen Blick auf entscheidende Fragen!

Wilco, jij ook bedankt voor je onuitputtelijke vrolijkheid! Ik weet niet hoe je het doet, maar je krijgt me altijd weer aan het lachen.

LUSV Badminton has played a major role in my social life and as such contributed essentially to my well-being. I would like to thank all members of the club for having had enough trust in me to make me their president. It was a pleasure being involved in what has become one of Leiden's most hospitable sport associations and to work together with Elien and Phát who were the best board I could have wished for. Elien, good luck in Melbourne! It's a shame that you won't be able to attend my thesis defense, but I hope to see you again some day. Phát, you will always remain a very special person for me and

although I might not see you for a while I will certainly not forget you. Within the club I found valuable friends and companions: Annelies, Bernadette, Bo, Eddy, Elien, Elise, Erik, Huda, Ion, Irene, Ivar, Lotte, Milou, Peter, Phát, Rahmat, Rayan, Remco, Richard, Sampo, Sierk, Suzanne, Thomas and Vera. Our legendary game nights, potlucks, walks on the beach and camp fires will always be a precious memory and I will miss you very much!

Sierk, Chris and Donald, thanks for your sympathies during different stages of the project and for sharing your philosophies with me.

Rod Suthers invited me to spend three months in his lab at Indiana University in Bloomington and the results of this collaboration are presented in chapter 3 of this thesis. I experienced a hearty welcome in his lab and a hospitable atmosphere. Amy, thanks for arranging housing for me and taking me horseback riding. I will never forget when we were driving to Lafayette and Bandit, a talking African grey parrot, started calling for help. Kori and Brent, thanks for accommodating me in the first week and everything else.

I also met a rather outstanding group of geocachers in Bloomington who never treated me as a stranger but always as one of them. Thanks to Jess for hooking me up with the group and special thanks to Stu, DJ and Mouse for our adventures in Hoosier National Forest, on (and in) Lake Monroe, on a ridge above the Ohio River and of course in the Great Smoky Mountains National Park. I hope to experience many more adventures with you in the future!

Caroline, you are much more than a colleague to me. You are a really good friend. Sometimes things are not going as smooth as we wish they would, but hang in there and I know you will make it!

Mama und Papa, ihr sagt zwar selbst, dass ihr nur die Hälfte von dem versteht was ich eigentlich mache, aber eurem Enthusiasmus hat das bisher zum Glück nicht geschadet. Danke fur ein stets offenes Ohr und eure nie endenden Sorgen um mich.

As John Howard Payne said in 1823: ‘There’s no place like home.’ For me, however, home is not a particular place. It’s the people I like to be with. It’s you. All of you.

Curriculum vitae, conference contributions & publications

Curriculum vitae

1983 Born on April 9 in Viersen, Germany.

2002 Graduation from high school (Albertus-Magnus-Gymnasium Viersen-Dülken, Germany): Abitur

2002-2007 MSc Biology, Heinrich-Heine-University Düsseldorf, Germany.

Specialisation: Cell biology, plant physiology and behavioural biology.

MSc thesis:
'Comparative investigations on the functional neuroanatomy of different breeds of domestic duck (*Anas platyrhynchos* f.d.)'
Supervisors: Prof. Dr. Hartmut Greven (Institute of Zoology, Heinrich-Heine-University Düsseldorf), Prof. Dr. Gerd Rehkämper, Dr. Julia Cnotka (Institute for Brain Research, Heinrich-Heine-University Düsseldorf).

2007-2011 PhD research project, Department of Behavioural Biology, Institute of Biology Leiden, Leiden University, The Netherlands.

Project title: Speech across Species: On the mechanistic fundamentals of vocal production and perception.

Supervisors: Prof. Dr. Carel ten Cate (Institute of Biology Leiden, Leiden University) and Dr. Gabriël J. L. Beckers (Max-Planck-Institute for Ornithology, Seewiesen, Germany).

During my PhD research project I spent three months (May-July 2010) in the lab of Prof. Dr. Roderick A. Suthers at Indiana University in Bloomington to collect cineradiographic data on vocalizing monk parakeets and speech-imitating African grey parrots. I supervised several BSc and MSc student projects and was involved in both theoretical and practical courses on behavioural biology.

Conference contributions

2010 Annual meeting of the Dutch Society for Behavioural Biology (NVG) in Soesterberg, The Netherlands. *Oral presentation.*

2010 Annual meeting of the Animal Behavior Society (ABS) in Williamsburg, Virginia. *Oral presentation.*

2010 8th International Conference on the Evolution of Language (Evolang 8) in Utrecht, The Netherlands. *Oral presentation.*

2010 Workshop 'Birdsong/animal communication and the evolution of speech', preceding Evolang 8. *Oral presentation.*

2009 Annual meeting of the Dutch Society for Behavioural Biology (NVG) in Dalfsen, The Netherlands. *Oral presentation.*

2009 IBL Symposium, Leiden University, The Netherlands. *Oral presentation.*

2009 31st International Ethological Conference (IEC) in Rennes, France. *Oral presentation.*

2008 Annual meeting of the Dutch Society for Behavioural Biology (NVG) in Dalfsen, The Netherlands. *Oral and poster presentation.*

2008 Vocal Communication in Birds and Mammals Conference in St. Andrews, Scotland. *Poster presentation.*

2007 Graduate meeting of the study group behavoural biology of the Ethological Association (Ethologische Gesellschaft e.V.) in cooperation with the German Zoological Association (Deutsche Zoologische Gesellschaft e.V.) in Göttingen, Germany. *Oral presentation.*

Publications

Published

Ohms, V. R., Snelderwaard, P. C., ten Cate, C. and Beckers, G. J. L. (2010). Vocal tract articulation in zebra finches. *PLoS ONE* 5: e11923.

Verzijden, M. N., Ripmeester, E. A. P., **Ohms, V. R.**, Snelderwaard, P. and Slabbekoorn, H. (2010). Immediate spectral flexibility in singing chiffchaffs during experimental exposure to highway noise. *Journal of Experimental Biology* 213: 2575-2581.

Ohms, V. R., Gill, A., van Heijningen, C. A. A., Beckers, G. J. L. and ten Cate, C. (2010). Zebra finches exhibit speaker-independent phonetic perception of human speech. *Proceedings of the Royal Society of London Series B- Biological Sciences* 277: 1003-1009.

Manuscripts

Ohms, V. R., Beckers, G. J. L., ten Cate, C. and Suthers, R. A. Vocal tract articulation revisited: the case of the monk parakeet.

Ohms, V. R., Escudero, P., Lammers, K. and ten Cate, C. Zebra finches and Dutch adults exhibit the same cue weighting bias in vowel perception.